In compliance with the
Canadian Privacy Legislation
some supporting forms
may have been removed from
this dissertation.

While these forms may be included
in the document page count,
their removal does not represent
any loss of content from the dissertation.

University of Alberta

STEPWISE CONDITIONAL TRANSFORMATION FOR MULTIVARIATE GEOSTATISTICAL SIMULATION

by

Oy Leuangthong ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

in

Mining Engineering

Department of Civil and Environmental Engineering

Edmonton, Alberta
Fall 2003

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisisitons et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-88013-3
Our file Notre référence
ISBN: 0-612-88013-3

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

Canada

University of Alberta

Library Release Form

Name of Author: Oy Leuangthong

Title of Thesis: Stepwise Conditional Transformation for Multivariate Geostatistical Simulation

Degree: Doctor of Philosophy

Year this Degree Granted: 2003

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Date: June 3, 2003

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Stepwise Conditional Transformation for Multivariate Geostatistical Simulation** submitted by Oy Leuangthong in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in *Mining Engineering.*

Dr. D.D. Tannant (Examiner)

Date: June 2, 2003

To Shawn, who has been there cheering for me every step of the way.
Your love, support and encouragement mean the world to me.

Also to my parents, Souk and Somluck, for the courage
and strength it took to provide a better life for us.

Abstract

Numerical models are most effective when they account for all sources and types of data. Real data exhibit complex multivariate features such as non-linear, constraint, and heteroscedastic relations. Current geostatistical simulation methods that allow for modeling of multiple variables rely on simple statistical models that are sometimes inappropriate or unable to account for realistic complexity in the multivariate relations.

This dissertation develops the stepwise conditional transformation technique for use as a pre- and post-processing tool for multivariate Gaussian simulation. The back transformation enforces reproduction of the original complex multivariate features. The methodology and underlying assumptions are explained. Several petroleum and mining examples are used to show features of the transformation and implementation details.

Application to the Red Dog zinc deposit showed an increase in profit from the simulation approach relative to the common practice of kriging. A comparative study of multivariate simulation using stepwise conditionally transformed variables against conventional simulation approaches is also shown. Further, the stepwise transform can also be used to account for multivariate features resulting from trend modeling.

Acknowledgements

Over the past four years, many doors have opened that I never would have anticipated. For this, I have Dr. Clayton Deutsch to thank. I am grateful for all his support, guidance and invaluable advice in my research as well as my professional endeavours.

I have also enjoyed the many hours of conversation with my fellow researchers in the Centre for Computational Geostatistics. The diversity in culture, background and especially opinions have made for an enjoyable graduate studies experience. Thanks to the many graduate students who have shared the CCG experience during this time. I would especially like to thank a couple of my fellow colleagues - Julian Ortiz and Karl Norrena - for their friendship and support over the last four years.

Financial support for this research was provided largely by the sponsors of the CCG, Killam Foundation and the Coal Mining Research. I would especially like to thank Teck Cominco Limited for permitting me to include my work on Red Dog in this dissertation. Their interest and support in my work has been greatly appreciated.

Finally and most importantly, I would like to thank my husband, Shawn, and my family for their unconditional love and support.

Table of Contents

1 Introduction 1

1.1 Problem Setting 1

1.2 An Introductory Example 3

1.3 Dissertation Outline 7

2 Concepts and Algorithms for Multivariate Simulation 8

2.1 Multivariate Geostatistics 8

2.1.1 Random Function Concept 8

2.1.2 Normal Score Transform 10

2.1.3 Measures of Spatial Variability 11

2.1.4 Models of Coregionalization 14

2.1.5 Volume Variance 20

2.1.6 Kriging 21

2.1.7 Simulation 25

2.2 Multivariate Statistics 32

2.2.1 Dimension Reduction Methods 33

2.2.2 Data Transformation Methods 35

2.2.3 Classification Techniques 36

3 Stepwise Conditional Transformation 41

3.1 Model of Coregionalization 46

3.2 Links to "Cloud" Transform / P-field 53

3.3 Remarks 55

4 Considerations for Real Data 56

4.1 Data Related Issues 56

4.2 Number of Data and Class Specification 59
4.3 Transformation in Presence of Sparse Data 62
4.4 Effect of Ordering 65
4.5 Remarks 71

5 Case Study: Multivariate Simulation of Red Dog Mine, Alaska, USA 74
5.1 Background 74
5.2 Available Data 75
5.3 Multivariate Geostatistical Simulation 75
  5.3.1 Validation of Simulation Models 81
  5.3.2 General Comments on Conditional Simulation Models 87
5.4 Validation with Additional Data 87
5.5 Potential Applications of Simulated Models 99
5.6 The Value of Simulation 114
5.7 Remarks 123

6 Comparison of Multivariate Cosimulation Algorithms 125
6.1 The Data 125
6.2 Cosimulation Algorithms 126
6.3 Comparison of Cosimulation Algorithms 137
6.4 Remarks 141

7 Stepwise Transformation for Geostatistical Modeling with a Trend 143
7.1 Background 143
7.2 Proposed Methodology 148
7.3 Application 148
  7.3.1 Mining Example 148
  7.3.2 Petroleum Example 153
7.4 Remarks 155

8 Concluding Remarks 159
8.1 Stepwise Conditional Transformation 159
8.2 Guidelines for Practical Application 160

8.3 Future Research in Multivariate Geostatistics 162

Bibliography **165**

A Program Details **171**

A.1 Kernel Smoothing 171

A.2 Stepwise Transformation: Typical Application 172

A.3 Stepwise Transformation: Kernel Smoothing 174

A.4 Back Transformation 174

B A Framework for Direct Sequential Simulation for Multiple Variables **176**

List of Figures

1.1 Schematic illustration of different bivariate distributions 3

1.2 Univariate and bivariate distributions of original data for Two Well. 4

1.3 Two Well simulation: one realization of porosity and cosimulated log permeability. 5

1.4 Comparison of cross plots before and after Gaussian simulation (in Gaussian space). 6

1.5 Comparison of cross plots before and after Gaussian simulation. . . . 6

2.1 Transformation of original data cumulative distribution (Z) to a standard normal cumulative distribution (Y). 11

2.2 Graphical interpretation of semi-variogram 12

2.3 Common negative semi-definite variogram models 13

2.4 Graphical interpretation of calibration parameter for Markov-Bayes updating 19

2.5 Schematic of volume variance. 20

2.6 ACE example 37

2.7 Schematic illustration of linear and quadratic discriminant analysis . 38

2.8 Clustering data into groups. 39

2.9 Hierarchical clustering: dendrogram or hierarchical tree. 40

3.1 Schematic illustration of stepwise conditional transformation of two variables. 42

3.2 Comparative illustration of cross plots for Mining data. 44

3.3 Comparative illustration of cross plots for Petroleum data. 45

3.4 Schematic illustration of two points separated by a distance **h**. . . . 47

3.5 Conditional distribution of the secondary variable, $Y_2(\mathbf{u})$ with conditional mean, $\mu_{2|1}(\mathbf{u})$ and variance, $\sigma_{2|1}(\mathbf{u})$. 47

3.6 Direct and cross variogram models for numerical example to investigate model of coregionalization. 51

3.7 Direct semivariogram after stepwise conditional transformation . . . 52

3.8 Checking input and output cross variograms after simulation. 53

3.9 Schematic illustration of cloud transform for petrophysical modeling. 54

4.1 Effect of outliers on stepwise conditional transformation. 57

4.2 Despiking for quantile transformation. 58

4.3 Despiking problem associated to class thresholds for stepwise transformation. 59

4.4 Effect of number of classes on correlation coefficient. 60

4.5 Differences in transformed secondary data due to class partition. . . 61

4.6 Dynamic class expansion to obtain minimum number of data for stepwise conditional transformation. 63

4.7 Effect of correlation and variance on bivariate kernel generated centered about a data pair. 64

4.8 Crossplot smoothing on small petroleum dataset consisting of only 27 samples. 65

4.9 Methodology to examine the effect of transformation ordering on spatial structures. 67

4.10 Effect of transformation ordering on Two Well Data. 69

4.11 Effect of ordering using East Texas data. 70

4.12 Example of independent simulation and back transformation of porosity and log permeability for Two Well Data. 73

5.1 Plan and cross sectional views of available data for Red Dog mine. . 76

5.2 Comparison of equally weighted Zn distribution and representative Zn distribution. 78

5.3 Calibration crossplot of Pb given Zn. 79

5.4 Comparison of equally weighted Pb distribution and representative Pb distribution. 79

5.5 Crossplot between stepwise conditionally transformed variables for Zn, Pb and Fe. 80
5.6 Variography for Zn. 82
5.7 Composites data reproduction for Zn. 84
5.8 Histogram reproduction for Zn: representative Zn histogram compared to simulated Zn histogram. 84
5.9 Reproduction of summary statistics for Zn: histogram of means and variances from multiple simulated realizations. 85
5.10 Variogram reproduction for Zn. 86
5.11 Comparison of multivariate features reproduction for Red Dog case study. 88
5.12 Simulated realizations of Zn at 12.5ft grid resolution. 89
5.13 Crossplot of BH data against nearest neighbour DH data for Zn, Pb, Fe and Ba over all eight rock types. 90
5.14 Simulated realizations of Zn at 12.5 x 12.5 x 25ft grid resolution, consistent with BH data. 91
5.15 Crossplots of BH data and E-type estimate from Conditional Simulation Models. 92
5.16 Schematic illustration of relation between expected correlation and estimation variance at BH spacing. 93
5.17 Configuration for determining average estimation variance. DH data are at 100 x 100ft spacing, BH data are at 10 x 12ft spacing. 94
5.18 Calculation of expected correlation for Zn within one rock type: histogram of estimation variance from kriging for 8 x 8 grid centered about a DH sample, and crossplot of BH data and E-type estimate. 95
5.19 Accuracy plot for E-type estimate of Zn data. 96
5.20 Accuracy plot for BH data and E-type estimate of simulation models for Zn, Pb, Fe and Ba. 97
5.21 Simulated realizations of Zn at 25ft grid resolution. 98
5.22 Comparison between existing long term model with E-type estimate from simulation at 25ft grid resolution. 100
5.23 Crossplots of Zn, Pb, Fe and Ba from E-type estimates and existing long-term model. 101

5.24 Probability maps to exceed a specific cutoff: $Zn > 5\%$, $Zn > 10\ \%$, $Zn > 25\ \%$, and $Ba > 7\ \%$. 103
5.25 Probability of ore map based on stockpile criteria. 104
5.26 Six realizations of the recovery models as calculated based on recovery functions provided by Teck Cominco. 106
5.27 Summary maps of the 40 recovery realizations: the minimum, average and maximum recovery at each location. 107
5.28 Uncertainty in the local recovery is shown for four arbitrarily chosen locations within the model. 108
5.29 Uncertainty in the global recovery based on all 40 realizations of recovery. 108
5.30 Six realizations of the resource models as calculated based on tonnage factors and recovery functions provided by Teck Cominco. 110
5.31 Summary maps of the 40 resource realizations: the minimum, average and maximum resource map at each location. 111
5.32 Uncertainty in the local resource is shown for four arbitrarily chosen locations within the model (same locations as shown in Figure 5.28). 112
5.33 Uncertainty in the global resource based on 40 realizations. 112
5.34 Illustration of application for short term planning: the volume of material associated to the planned production for one month, and the uncertainty in the resource available. 113
5.35 Zn recovery function developed for comparison of kriging and simulation. 115
5.36 Discount function on Zn recovery due to %Ba content. 115
5.37 Discount function on Zn recovery due to %Fe content. 116
5.38 Location map of reference BH data and sampled BH data for use in comparing model approaches. 117
5.39 Variograms of direct space data for use in ordinary kriging approach. 118
5.40 Variograms of stepwise conditional scores for use in simulation approach. 119
5.41 Comparison of kriged model and one realization from simulation. . . 120
5.42 Comparison of true profit map at data locations and the profit map for ore/waste classification from kriging and simulation. 121

5.43 Comparison of true ore/waste classification and the classification from kriging and simulation at data locations. 122
5.44 Ore/Waste classification summary of simulation and kriging relative to true ore/waste classification. 123

6.1 Location map of available hard data. 126
6.2 Histogram of porosity, permeability and seismic data. 126
6.3 Cross plot of Amoco data. 127
6.4 Cross plot of normal scores data for Amoco data. 128
6.5 Direct and cross variograms in horizontal direction for Amoco data. 129
6.6 Direct and cross variograms in vertical direction for Amoco data. . . 129
6.7 Cross plot of stepwise conditionally transformed scores. 131
6.8 Direct variograms for stepwise conditional scores of porosity and permeability. 131
6.9 Indicator variograms for porosity for each of the nine thresholds. . . 133
6.10 Indicator vertical variograms for porosity for each of the nine thresholds. 134
6.11 Indicator variograms for permeability for each of the seven thresholds. 135
6.12 Indicator vertical variograms for permeability for each of the seven thresholds. 136
6.13 Comparison of realizations from different cosimulation techniques. . 138
6.14 Comparison of flow results from different cosimulation techniques. . 139
6.15 Comparison of bivariate distribution reproduction from different cosimulation methods. 140
6.16 Comparison of normal scores cross plots for different Gaussian cosimulation approaches. 142

7.1 Example of vertical trend as indicated by two well logs. 145
7.2 Example of porosity log and corresponding vertical variogram showing existence of a vertical trend. 145
7.3 Example of heteroscedastic variance of residuals, and linear constraint on residuals. 147
7.4 Location map of available drillholes and crossplot of elevation vs. Cu to illustrate 1-D trend in the vertical direction. 149

7.5 Location map of vertically averaged Cu data, and resulting kriged map using this data. 149
7.6 Required 1D trend and 2D trend for integration to obtain 3D Cu trend model. 150
7.7 Linear constraint on residual Cu values due to trend component. . . 151
7.8 Crossplot of Cu trend vs. normal scores of Cu residual. 151
7.9 Histogram of transformed Cu residual and crossplot of Cu trend vs. the conditionally transformed residual components. 152
7.10 Comparison of trend model and simulated realization of Cu, after adding the trend back to the simulated residuals. 152
7.11 Comparison of declustered Cu distribution with the first realization of simulated Cu, after adding the trend component to the simulated residuals. 153
7.12 Comparison of original Cu trend-residual crossplot and modeled Cu trend - simulated residual crossplot. 153
7.13 Crossplot of the modeled Cu trend vs. simulated Cu residuals from applying the conventional normal score transform. 154
7.14 Location map of available wells and crossplot of elevation vs. core porosity to illustrate 1-D trend in the vertical direction. 154
7.15 Location map of vertically averaged porosity data and resulting areal trend map using this data. 155
7.16 Required 1D vertical trend and 2D areal trend used to construct a 3D porosity trend. 156
7.17 Relation between residual porosity values and collocated trend component. 156
7.18 Histogram of stepwise conditionally transformed (SC) porosity residual, and crossplot of the trend vs. the conditionally transformed residual components. 157
7.19 Comparison of porosity trend model and a realization of simulated porosity. 157
7.20 Comparison of declustered porosity distribution with the first realization of simulated porosity. 157

7.21 Comparison of original porosity trend-residual crossplot and modeled trend - simulated residual crossplot. 158

A.1 Parameters for **scatsmth_k**. 172
A.2 Parameters for **stepcon**. 173
A.3 Parameters for **stepcon_k**. 174
A.4 Parameters for **backstep**. 175

B.1 Example: 2-D map of core and seismic data for generalized cokriging. 180
B.2 Results from applying iterative updating approach to a global standard bivariate Gaussian distribution 184
B.3 Schematic illustration of combining subsets of bivariate distributions 186

List of Tables

5.1 Red Dog model coordinate limits. 75

5.2 Transformation ordering for stepwise conditional transformation. . . 77

5.3 Summary table for determining expected correlation coefficient for one rock type. 94

5.4 Correlation between BH [illegible]a and Long Term Model values. 95

5.5 Summary of correlation coefficients from all comparisons: BH to DH, BH to Long Term (LT) model, BH to E-type estimate, and Long Term model to E-type estimate. 99

6.1 Table of variance contribution (cc) of each nested structure in the six variograms constituting an LMC model for full cokriging. 128

6.2 Table of calibration factors between the hard local data and available soft data. 132

6.3 Summary of p_{10}, p_{50} and p_{90} effective permeability in X direction from each of the cosimulation methods employed. 139

Chapter 1

Introduction

Geostatistics is a relatively new and rapidly growing area in engineering, the earth sciences, and applied mathematics. The field is devoted to the application of statistical techniques in the study of spatially variable phenomena. Although geostatistics was first developed to improve ore reserve estimation in a mining context, it has grown in application to many other areas of the earth sciences.

A significant advantage to the application of geostatistics is the ability to model uncertainty. Simulation results in a set of realizations that honour the data, the global histogram, and the spatial correlation. The difference between realizations provides a measure of uncertainty. An assessment of uncertainty makes it possible to evaluate risk more accurately and make better-informed decisions.

In the case of multiple variables, multivariate geostatistics provides the modeling tools to account for additional information from different types/sources of data. Accounting for the relationship between multiple variables can greatly improve the representativity of these numerical models. Risk assessment and decision-making based on these multivariate models will be better.

1.1 Problem Setting

The construction of geologically realistic models depends on the integration of data of different types, sources, and volumes of measurement. For resource characterization, available data may consist of field data, analogue data and expert knowledge. For example in petroleum reservoirs, field data may include core samples, well logs, and seismic data. In a mining context, several mineral grades may be available from drill holes and blast holes. In all cases, the numerical models should account for field data and conform to the geologist's interpretation of the depositional environment.

Geostatistical simulation permits the construction of realizations that honour the data, histogram and spatial variability. Classical techniques require that conditional distributions of the attribute of interest be defined at each location $\mathbf{u}$ within the domain A. In the univariate case, the conditional distribution at location $\mathbf{u}$ is based on (or conditional to) correlated data of the same type. Similarly, the multivariate case requires that the conditional distribution at each location account for correlated data of *multiple* types. The task of conditioning local distributions to multiple variables is a difficult problem. Differences in distribution, sampling density, spatial continuity, and size scales of the data types complicate the inference of this distribution.

Local conditional distributions are determined by assigning weights to nearby data. These weights are traditionally obtained by solving the kriging system of equations. Kriging accounts for data redundancy, closeness and correlation. The relationship between the data is described by a model of coregionalization. The only practical model that is mathematically consistent is the linear model of coregionalization (LMC) [17, 28, 40]. A simplification of the LMC is achieved in the Markov model for collocated data [4, 28, 81]. Based on the same theoretical foundation as the LMC and Markov model, the Markov-Bayes model was developed specifically for indicator simulation [42].

In conventional simulation, the role of kriging is to determine the local conditional distribution from which a simulated value is drawn. The most common simulation approach is Gaussian simulation, which is based on the simplest multivariate distribution - the Gaussian (or normal) distribution. In the Gaussian framework, the kriged estimate is exactly the conditional expectation of a Gaussian distribution, and the kriging estimation variance is exactly the conditional variance of this same Gaussian distribution [15, 28, 40]. Thus in this context, kriging based on a model of coregionalization provides the two essential parameters to define the local Gaussian distribution: the mean and variance.

In the presence of two or more variables, the conventional procedure for Gaussian simulation is to transform each variable to a Gaussian distribution one at a time. This ensures that each variable is univariate Gaussian; however, the multivariate distributions (of two or more variables at a time) are not explicitly transformed to be multivariate Gaussian although they are assumed to be multiGaussian. Real multivariate distributions may show non-Gaussian features such as non-linear, heteroscedastic, and/or mineralogically constrained features (see Figure 1.1). In these instances, simulating within the conventional Gaussian framework may not correctly reproduce the spatial variability of the phenomena.

The main objective of multivariate geostatistics is to improve the accounting of additional information to obtain models that are closer to the truth. Poorly reproducing multivariate relations is a serious problem.

The objective of this research is to address the problem of integrating multiple variables in geostatistical simulation. The research aims to improve upon the conventional multivariate Gaussian simulation.

The first development is to improve conventional Gaussian cosimulation by ensuring adherence to multiGaussian assumptions. The stepwise conditional transformation is a robust technique that ensures multivariate Gaussianity of collocated transformed variables. The method is robust in handling problematic multivariate distributions, and is effective as a pre- and post-processing algorithm to Gaussian simulation. Theoretical and practical implementation details of this technique will be explored.

The practicality of the transformation method in various applications will be examined. These applications include: the multivariate simulation of a real, complex mineral deposit; a comparative study with more traditional multivariate simulation

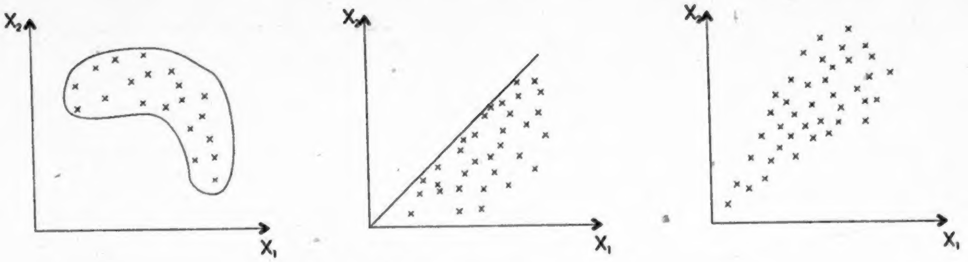


Figure 1.1: Schematic illustration of different bivariate distributions: non-linear (left), constraint (centre) and heteroscedastic (right).

techniques; and application of the transformation to geostatistical simulation with a trend model.

The results of this research affect the characterization of natural resources. Accounting for multiple data types will reduce the uncertainty in the models and improve characterization of heterogeneities. Petroleum, mining, environmental, forestry, and agricultural industries will benefit from numerical models that are more realistic.

1.2 An Introductory Example

A petroleum dataset referred to as the "Two Well" data is considered as a specific example to illustrate (1) the conventional Gaussian cosimulation of two model variables, and (2) the inherent assumption of multiGaussianity in the cosimulation process.

The two variables of interest are porosity and permeability; Figure 1.2 shows the univariate and bivariate distributions of the original variables and also the bivariate distribution of the normal scores of both variables, after normal score transformation. Prior to simulation, the conventional modeling approach entails that both the porosity and log permeability must be transformed to normal scores and the variography determined for these normal scores. Application of Gaussian cosimulation inherently assumes that the bivariate and all higher order distributions are also Gaussian.

Note that in the original bivariate distribution, the region of high porosity values and low log permeability values (that is, the region approximately defined by $\phi > 0.18$ and $logk < 2.5$) appears to have two constraints. After normal scores transformation of both variables, the bivariate distribution of the normal scores also exhibits a strange non-linear constraint.

Sequential Gaussian simulation was performed on porosity, and collocated cosimulation was performed for log permeability. Figure 1.3 shows a sample realization for porosity and the corresponding cosimulated realization for log permeability. For comparison, the original bivariate distribution is provided next to the crossplot of the simulated porosity and log permeability. The constraints that were noted in the

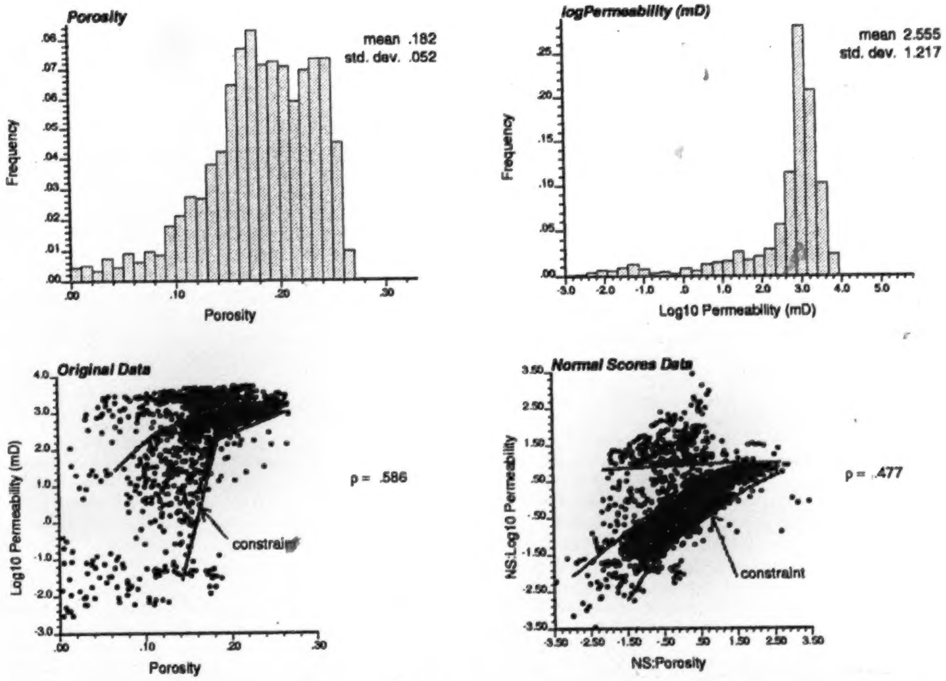


Figure 1.2: Univariate and bivariate distributions of original data. The top two histograms give the distribution of the original porosity (top left) and log permeability (top right); the bottom crossplots show the relation of the original data (bottom left) and the relations between the corresponding normal scores (bottom right).

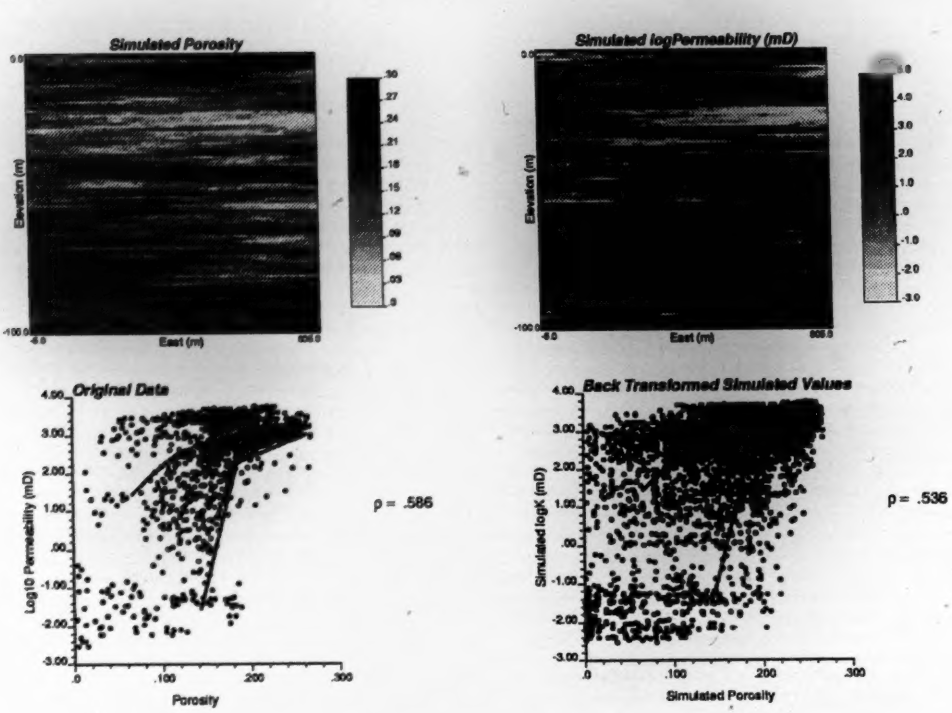


Figure 1.3: One realization of porosity (top left) and cosimulated log permeability (top right), using normal score transformation. The bottom figures show the cross plot of the original data (bottom left) and the cross plot of the cosimulated values (bottom right).

original data cross plot are also shown on the realization crossplot; the constraints are not apparent in the realization crossplot. Note that the density of simulated values on the crossplot makes it difficult to discern whether the top constraint is reproduced; however, it is clear that the lower right constraint is not honoured. The correlation coefficient is approximately reproduced.

Although the cosimulated results in original units reproduce the correlation between the input variables (Figure 1.3), the cross plot of the cosimulated values in Gaussian space does not reproduce the same bivariate relations between the data normal scores. This is shown in Figure 1.4. Slight differences in the correlation coefficient (for both original and Gaussian space crossplots) are attributed to statistical fluctuations. Departures of the conditioning data from the bivariate Gaussian assumption leads to a poor reproduction of the bivariate relations between the two model variables. This poor reproduction of the relationship between porosity and permeability may have serious consequences in exploration and/or production drilling. In the regions of the reservoir where the simulated values (for both porosity and permeability) fall outside the bounds of the constraints in Figure 1.3, drilling with the expectation of finding relatively high porosity and permeability may prove to be a costly venture.

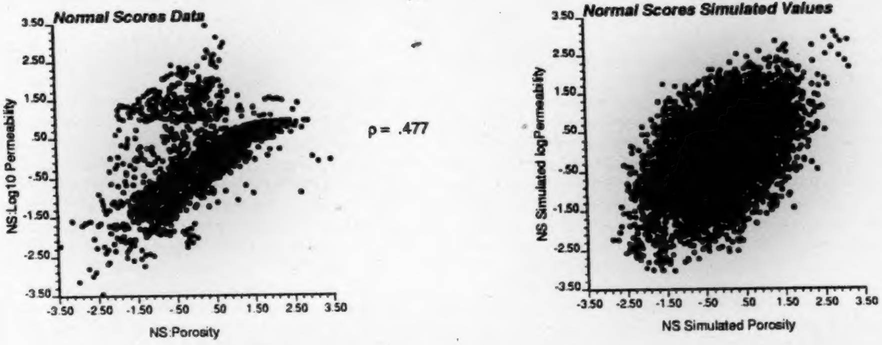


Figure 1.4: Comparison of cross plots generated by normally transforming the model variables (left) and the normal score values of the cosimulated results (right).

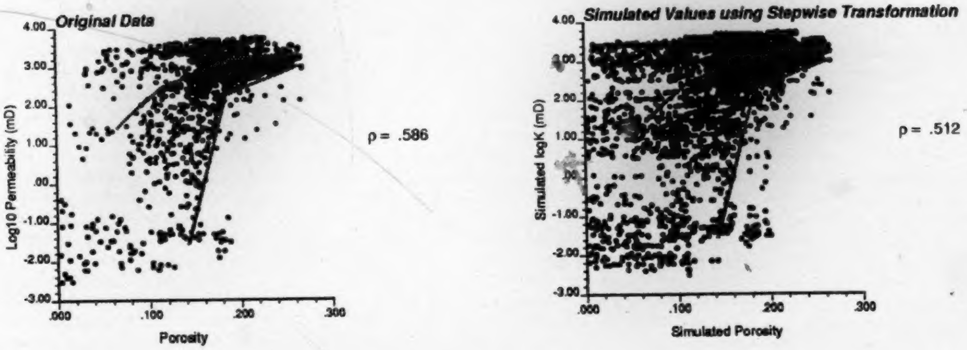


Figure 1.5: Comparison of the original data cross plot with the simulated values (in original space) using stepwise conditionally transformed variables.

As a preview to the effectiveness of the stepwise conditional transformation in honouring the multivariate data distribution, Figure 1.5 shows the crossplot of the original data alongside the crossplot of the simulated results when the stepwise transformation was applied. As with the simulated values using the conventional normal score transform, the density of simulated values on the back transformed results using stepwise transformed variables makes it difficult to see if the top constraint is reproduced. It is clear that the bottom constraint is better reproduced by applying the stepwise conditional transform than the normal score transform. This improved reproduction of the multivariate distribution yields models of porosity and permeability that are closer to the reality (based on the available data). As a result, risk analysis and decision-making based on these models are better in reducing potential economic loss.

1.3 Dissertation Outline

Chapter 2 contains a literature review including (1) a review of fundamental geostatistical concepts required to understand the current state of multivariate geostatistics, and (2) an overview of multivariate statistical techniques that have potential for future applications in geostatistics.

Chapter 3 introduces the stepwise conditional transform as a multivariate Gaussian transformation technique to improve Gaussian simulation. Practical issues to implementation are addressed in Chapter 4.

Chapter 5 presents a real application of the transformation for multivariate simulation of Red Dog mine in Alaska, USA. A comparison to the conventional practice of kriging is also shown in this chapter. A comparison of the stepwise transformation against other conventional multivariate simulation approaches is shown in Chapter 6.

Chapter 7 presents an application of the stepwise conditional transformation for geostatistical simulation to account for a trend. This application of the transformation shows the widespread applicability of the technique.

Chapter 8 concludes this dissertation with some final comments regarding the transformation, its place in multivariate geostatistics, and other future works yet to be explored in this area.

Appendix A provides specific details about the prototype programs developed to implement the stepwise conditional transformation. Appendix B consists of preliminary research results in the development of a direct sequential cosimulation framework.

Chapter 2

Concepts and Algorithms for Multivariate Simulation

2.1 Multivariate Geostatistics

Geostatistics is the practical application of probability theory to the modeling of natural phenomena. Since its birth in the 1960s, many different parametric and non-parametric geostatistical techniques have emerged. These algorithms all rely on the concept of a random function and the definition of a spatial distribution corresponding to that random function.

Multivariate datasets are common in geostatistics. Conventional geostatistical tools make assumptions regarding data distributions, stationarity of population statistics, and linear correlation between variables. Unfortunately, most sample datasets do not conform to all these assumptions.

2.1.1 Random Function Concept

A random variable (RV), Z (denoted by upper case letter), is a variable that can take a series of possible values as characterized by a probability density function (pdf) or equivalently a cumulative distribution function (cdf). Spatial dependence of the RV is denoted by $Z(\mathbf{u})$, where $\mathbf{u}$ is a location within the domain A. A particular outcome at some location $\mathbf{u}$ is denoted by $z(\mathbf{u})$ (lower case letter). The cdf characterizing the uncertainty in a RV $Z(\mathbf{u})$ is:

$$F_Z(\mathbf{u};z) \quad = \quad Prob\{Z(\mathbf{u}) \leq z\} \tag{2.1}$$

A random function (RF) is the set of dependent random variables, $\{Z(\mathbf{u}), \mathbf{u} \in A\}$. The RF is defined by its spatial law [17, 21, 28].

$$F_{Z(\mathbf{u}_1),\dots,Z(\mathbf{u}_N)}(z(\mathbf{u}_1),\dots,z(\mathbf{u}_N)) = Prob\{Z(\mathbf{u}_1) \leq z(\mathbf{u}_1),\dots,Z(\mathbf{u}_N) \leq z(\mathbf{u}_N)\}, \quad \forall \mathbf{u}_\alpha \in A, \alpha = 1,\dots,N \tag{2.2}$$

where N is the number of locations in domain A.

Notation

The mathematical notation adopted throughout the literature review is introduced.

For the univariate case or the case where we are only interested in the primary variable, the random variable is defined as

$$Z(\mathbf{u}_\alpha), \alpha = 1, \ldots, N$$

α = index that specifies location in domain A,
N = total number of locations in domain A

Note that N can be potentially very large.

In the multivariate case, an additional subscript index is required to specify the data type, that is,

$$Z_i(\mathbf{u}_\alpha), i = 1, \ldots, P, \alpha = 1, \ldots, N$$

where

i = index that specifies the data type,
P = total number of different data types,

A particular outcome of the random variable is similarly defined except with z (lower case) replacing Z (upper case). Furthermore, where reference is made to some arbitrary location in the domain A (as in theory development in the following sections), then the subscript α will be dropped for simplicity. The number of data locations is n where $n < N$.

The first and second order moments of the RF will be denoted as:

1. Mean:
    - For the univariate case,

$$E\{Z(\mathbf{u})\} = \mu(\mathbf{u})$$

    - For the multivariate case,

$$E\{Z_i(\mathbf{u})\} = \mu_i(\mathbf{u})$$

2. Covariance:
    - For the univariate case,

$$Cov\{Z(\mathbf{u}) \cdot Z(\mathbf{u}+\mathbf{h})\} = C(\mathbf{u}, \mathbf{u}+\mathbf{h})$$

    - For the multivariate case,

$$Cov\{Z_i(\mathbf{u}) \cdot Z_j(\mathbf{u}+\mathbf{h})\} = C_{ij}(\mathbf{u}, \mathbf{u}+\mathbf{h}),\ i, j = 1, \ldots, P$$

where $\mathbf{h}$ is a distance vector from location $\mathbf{u}$.

Stationarity. The concept of stationarity is critical to inference. Stationarity allows the geostatistician to extend his exploratory statistical analysis from limited data to the entire domain of interest.

Stationarity is a decision that assumes invariance of the multivariate cdf over the domain, that is

$$F_{Z(\mathbf{u}_1),\dots,Z(\mathbf{u}_N)}(z(\mathbf{u}_1),\dots,z(\mathbf{u}_N)) = F_{Z(\mathbf{u}_1+\mathbf{h}),\dots,Z(\mathbf{u}_N+\mathbf{h})}(z(\mathbf{u}_1+\mathbf{h}),\dots,z(\mathbf{u}_N+\mathbf{h})), \forall \mathbf{h}$$

Invariance of the multivariate cdf implies invariance of all lower level (i.e. univariate, bivariate, $\dots,(k-1)$-variate) distributions and all lower level moments [21].

First order stationarity amounts to invariance of the first order moment or the mean; second order stationarity implies invariance of the second order moment or the covariance. For practical purposes, second order stationarity is often sufficient for geostatistical inference, i.e. $E\{Z_i(\mathbf{u})\} = \mu_i$ and $Cov\{Z_i(\mathbf{u}) \cdot Z_j(\mathbf{u}+\mathbf{h})\} = C_{ij}(\mathbf{h}), \forall i, j, \mathbf{h}$ and $\mathbf{u} \in A$.

Ergodicity. Another concept that is closely related to stationarity is that of ergodicity. Geostatistical simulation results in a simulated value for each RV within the domain, A; the set of these simulated values over the domain is often referred to as a realization of the RF model. If this domain is sufficiently large, then the ergodic theorem states that the mean of each realization approximates the expected value of the RF [55]. For practical purposes, an ergodic domain should be large *relative* to the range of the covariance structure of the RF, $\{Z(\mathbf{u}), \mathbf{u} \in A\}$ [21]. Minor differences between the realization statistics and the model statistics is referred to as ergodic fluctuations [28].

2.1.2 Normal Score Transform

The normal or Gaussian distribution is characterized by two parameters - its mean μ and variance σ^2. The probability density function that produces its characteristic bell shape is:

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right] \tag{2.3}$$

A standard normal distribution refers to a mean of zero and a unit variance. The corresponding pdf is:

$$g(y) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{y^2}{2}\right) \tag{2.4}$$

The normal distribution is the limit distribution for the Central Limit Theorem. For a multiGaussian RF, the linear combination of two or more RVs is normally distributed [35]. In fact, the sum of independent RVs following any distribution tends toward a normal distribution [35]. This result implies great simplicity for simulation and is the main reason why Gaussian approaches are commonly used. However, most earth sciences variables are not normally distributed. In order to

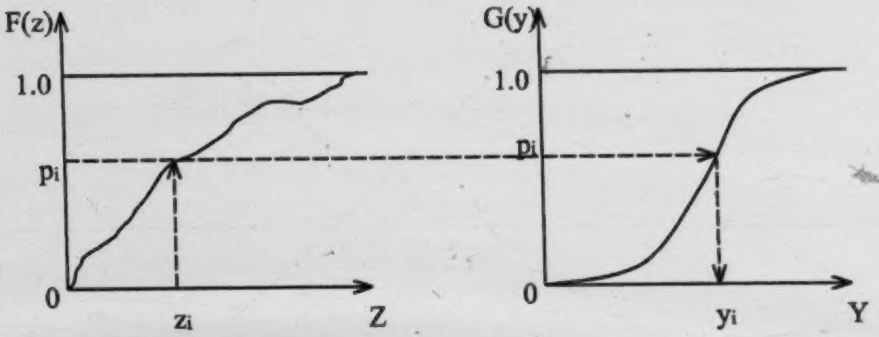


Figure 2.1: Transformation of original data cumulative distribution (Z) to a standard normal cumulative distribution (Y).

apply the Gaussian approach, the variables must first be transformed to normal space.

The basic steps in the common "graphical" or "quantile" transformation process are described below and illustrated in Figure 2.1:

1. The sample cumulative distribution function of the original data variable, Z, must be calculated. The transformation from one distribution to another is accomplished based on the probability associated to each data value.
2. For each sample data, $z_i, i = 1, \ldots, n$, the corresponding cumulative probability is identified, $p_i, i = 1, \ldots, n$. Once determined, the normal score value, $y_i, i = 1, \ldots, n$, corresponding to each probability is found:

$$y_i = G^{-1}[F(z_i)] = G^{-1}(p_i) \tag{2.5}$$

This transformation is commonly referred to as the normal score transform.

Once transformation is complete, the transformed variables (Y_i) are used in subsequent geostatistical calculations. Alternative transformation procedures such as fitting Hermite polynomials could be considered [13, 40]. The results of simulation must then be back transformed to original units. The back transformation follows a similar procedure, except the direction of transformation in Figure 2.1 is reversed.

2.1.3 Measures of Spatial Variability

The two most common measures used to describe the spatial continuity of a phenomena are the variogram and the covariance functions.

The variogram, $2\gamma(\mathbf{h})$, is a two-point statistic used as a measure of *dissimilarity* between two random variables separated by a Euclidean distance, $\mathbf{h}$,

$$2\gamma(\mathbf{h}) = E\{[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]^2\} \tag{2.6}$$

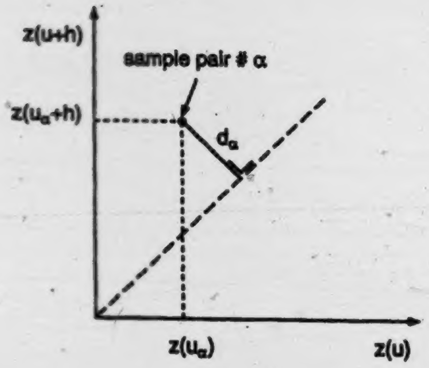


Figure 2.2: Semi-variogram interpretation: moment of inertia about the 45° line. Source: Goovaerts (1997)

where $Z(\mathbf{u})$ is a spatially distributed RV. In practice, the experimental variogram (denoted with by the symbol $\hat{\gamma}$) is calculated as the average squared difference between data approximately separated by $\mathbf{h}$:

$$2\hat{\gamma}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]^2 \tag{2.7}$$

where $N(\mathbf{h})$ is the number of pairs of data approximately separated by $\mathbf{h}$. The symbol $\gamma(\mathbf{h})$ is known as the semi-variogram and is simply half the variogram. The semi-variogram can be interpreted as the "moment of inertia of the scattergram around the 45° line" [28, 39] (See Figure 2.2).

Alternatively, the covariance function is a measure of *similarity* between data pairs:

$$C(\mathbf{h}) = E\{[Z(\mathbf{u}) \cdot Z(\mathbf{u}+\mathbf{h})]\} - \mu(\mathbf{u}) \cdot \mu(\mathbf{u}+\mathbf{h}) \tag{2.8}$$

Note that the covariance of a single random variable at $\mathbf{h} = 0$ is the variance of that random variable, i.e., $C(\mathbf{0}) = \sigma_Z^2$.

There are some interesting features that are common to both measures:

- Under stationarity, the semivariogram, covariance and variance are related:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}) \tag{2.9}$$

- For spatial inference, the covariance function must ensure that the variance of linear estimators are positive. This amounts to requiring that the covariance

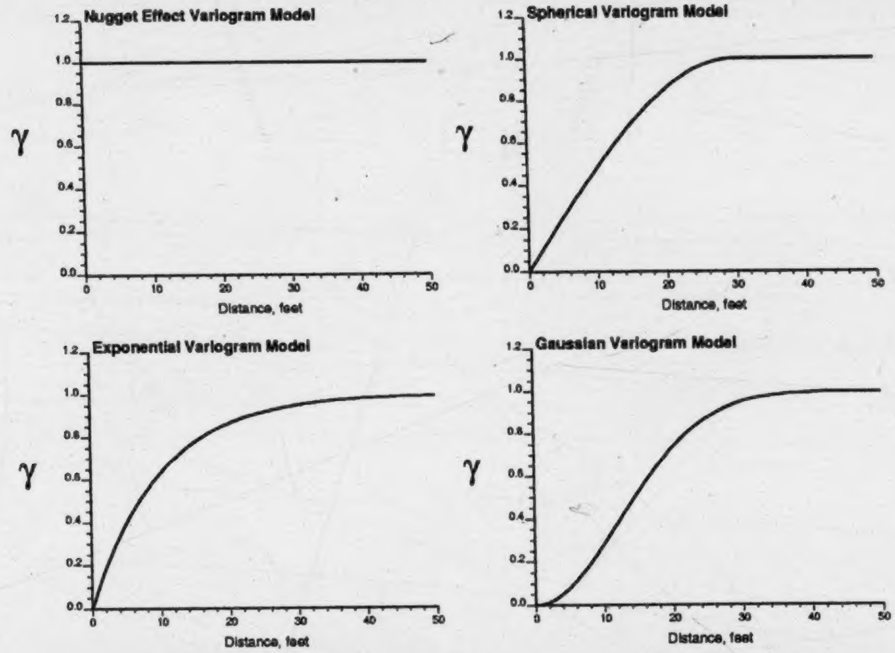


Figure 2.3: Common negative semi-definite variogram models: nugget effect (top left), spherical (top right), exponential (bottom left) and Gaussian (bottom right). Source: MIN E 612 Course Notes (1999)

function be positive semi-definite or the variogram function be negative semi-definite. Several variogram models are known to satisfy this constraint and are commonly used in variogram modeling. These include the nugget effect, spherical, exponential, and Gaussian models (See Figure 2.3).

For two or more variables, the spatial relationship between pairs of random variables is also required. For this purpose, the cross-variogram (Equation 2.10), cross-covariance (Equation 2.11) and their corresponding relationship (Equation 2.12) are given below:

$$2\gamma_{ij}(\mathbf{h}) = E\{[Z_i(\mathbf{u}) - Z_i(\mathbf{u}+\mathbf{h})][Z_j(\mathbf{u}) - Z_j(\mathbf{u}+\mathbf{h})]\} \tag{2.10}$$

$$C_{ij}(\mathbf{h}) = E\{[Z_i(\mathbf{u}) - \mu_i(\mathbf{u})][Z_j(\mathbf{u}+\mathbf{h}) - \mu_j(\mathbf{u}+\mathbf{h})]\} \tag{2.11}$$

$$C_{ij}(\mathbf{h}) = C_{ij}(\mathbf{0}) - \gamma_{ij}(\mathbf{h}) \tag{2.12}$$

Note that the cross covariance function may not necessarily be symmetric, that is, $C_{ij}(\mathbf{h}) \neq C_{ji}(\mathbf{h})$. This is commonly referred to as the lag effect [28, 40]. An example of such a case is provided by Journel and Huijbregts, wherein the rich grades of Pb lag behind those of Zn in a specific direction [40].

For standardized variables, C_{ij} in Equations 2.11 and 2.12 can be replaced by ρ_{ij}, the correlation coefficient between Z_i and Z_j.

Determination of the cross variogram or cross covariance is also based on experimental data. As a result, the modeling of direct and cross covariance must be modeled in a mathematically consistent manner, respecting physical laws and ensuring positive finite variances [17]. Models of coregionalization model the spatial continuity of two or more variables.

2.1.4 Models of Coregionalization

Consider a multivariate dataset consisting of P types of data. Explicit characterization of the spatial relationship between the P variables requires a matrix of covariance functions $C_{ij}(\mathbf{h}), i, j = 1, \dots, P$:

$$C(\mathbf{h}) = \begin{bmatrix} C_{11}(\mathbf{h}) & \cdots & C_{1P}(\mathbf{h}) \\ \vdots & \ddots & \vdots \\ C_{P1}(\mathbf{h}) & \cdots & C_{PP}(\mathbf{h}) \end{bmatrix}, \forall \mathbf{h}$$

This covariance matrix is often assumed to be symmetric (i.e. $C_{ij}(\mathbf{h}) = C_{ji}(\mathbf{h})$). Note that each element, $C_{ij}(\mathbf{h})$, represents an $n_i \times n_j$ matrix of covariances where n_i is the number of data for the i^{th} variable. To ensure that all variances are positive, the covariance matrix must be positive semi-definite, i.e. all leading principal minor determinants of order k must be non-negative, $k = 1, \dots, P$, i.e.

$$\det C(\mathbf{h}) = \sum_{i=1}^{P} (-1)^{i+j} C_{ij}(\mathbf{h}) \det U_{ij}(\mathbf{h}) \geq 0$$

where $\det U_{ij}(\mathbf{h})$ is a minor determinant of order $P-1$ of the $P \times P$ covariance matrix C, with the indices i and j denoting the row and column of C removed in order to form the $(P-1) \times (P-1)$ matrix [30, 61]. For the simple case of second order covariance matrix, the positive semi-definite constraint requires that

$$C_{ii}(\mathbf{h}) C_{jj}(\mathbf{h}) \geq C_{ij}(\mathbf{h}) C_{ji}(\mathbf{h}), \forall i, j, \mathbf{h}$$

Linear Model of Coregionalization (LMC)

Consider P stationary random functions, $Z = \{Z_1, \dots, Z_P\}$. Suppose that each random function $Z_i, i = 1, \dots, P$ can be expressed as a linear combination of K independent second-order stationary random functions, $Y_k, k = 1, \dots, K$, each with zero mean and covariance function $C_k(\mathbf{h})$:

$$Z_i(\mathbf{u}) = \sum_{k=1}^{K} a_{ik} Y_k(\mathbf{u}) + \mu_i \tag{2.13}$$

where

$$\begin{aligned} E\{Z_i(\mathbf{u})\} &= \mu_i \\ E\{Y_k(\mathbf{u})\} &= 0 \\ C\{Y_k(\mathbf{u}), Y_{k'}(\mathbf{u}+\mathbf{h})\} &= C_k(\mathbf{h}), \text{ if } k = k' \\ &= 0, \text{ otherwise} \end{aligned} \tag{2.14}$$

Note that the RFs $Y_k, k = 1, \ldots, K$ are underlying and unknown. If the RFs $Y_k, k = 1, \ldots, K$ are grouped by those RFs Y_k with the same direct covariances $C_k(\mathbf{h})$, then Equation 2.13 can be written as:

$$Z_i(\mathbf{u}) = \sum_{l=0}^{L} \sum_{k=1}^{n_l} a_{ik}^l Y_k^l(\mathbf{u}) + \mu_i \tag{2.15}$$

with

$$\begin{aligned} C\{Y_k^l(\mathbf{u}), Y_{k'}^{l'}(\mathbf{u}+\mathbf{h})\} &= C^l(\mathbf{h}), \text{ if } k = k' \text{ and } l = l' \\ &= 0, \text{ otherwise} \end{aligned} \tag{2.16}$$

where $L+1$ is the number of groups with different direct covariances, and n_l is the number of RFs with the same covariance in group l. Based on Equation 2.15, the cross covariance of two RVs $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u}+\mathbf{h})$ is

$$\begin{aligned} C_{ij}(\mathbf{h}) &= E\left\{ \left(\sum_{l=0}^{L} \sum_{k=1}^{n_l} a_{ik}^l Y_k^l(\mathbf{u}) \right) \left(\sum_{l'=0}^{L} \sum_{k'=1}^{n_{l'}} a_{jk'}^{l'} Y_{k'}^{l'}(\mathbf{u}+\mathbf{h}) \right) \right\} \\ &= \sum_{l=0}^{L} \sum_{l'=0}^{L} \sum_{k=1}^{n_l} \sum_{k'=1}^{n_{l'}} a_{ik}^l a_{jk'}^{l'} C\{Y_k^l(\mathbf{u}) Y_{k'}^{l'}(\mathbf{u}+\mathbf{h})\} \end{aligned} \tag{2.17}$$

Using the simplified covariance in Equation 2.16 simplifies Equation 2.17 to

$$\begin{aligned} C_{ij}(\mathbf{h}) &= \sum_{l=0}^{L} \sum_{k=1}^{n_l} a_{ik}^l a_{jk}^l C\{Y_k^l(\mathbf{u}) Y_k^l(\mathbf{u}+\mathbf{h})\} \\ &= \sum_{l=0}^{L} \sum_{k=1}^{n_l} a_{ik}^l a_{jk}^l C^l(\mathbf{h}) \end{aligned} \tag{2.18}$$

From Equation 2.18, the sill of the l^{th} covariance structure, $C^l(\mathbf{h})$, is given by $\sum_{k=1}^{n_l} a_{ik}^l a_{jk}^l$. Defining $b_{ij}^l, i = 1, \ldots, P, j = 1, \ldots, P$ such that

$$b_{ij}^l = \sum_{k=1}^{n_l} a_{ik}^l a_{jk}^l$$

simplifies Equation 2.18 to

$$C_{ij}(\mathbf{h}) = \sum_{l=0}^{L} b_{ij}^{l} C^{l}(\mathbf{h}) \tag{2.19}$$

It only remains to determine $C^{l}(\mathbf{h}), l = 0, \ldots, L$ and the $(L+1) \cdot P^2$ parameters b_{ij}^{l} so that covariances are jointly positive definite. If the covariance models $C^{l}(\mathbf{h}), l = 0, \ldots, L$ are chosen to be known positive semi-definite models, this amounts to requiring that the $L+1$ matrices of b_{ij}^{l} coefficients are also positive semi-definite [17, 28, 33, 40]. For 2 variables, this constraint requires that

$$b_{ii}^{l} \cdot b_{jj}^{l} \geq b_{ij}^{l} \cdot b_{ji}^{l}, \forall i, j, l$$

In practice, this requires that for P variables, a total of $P(P+1)/2$ licit covariances are required to be modeled simultaneously to ensure positive semi-definiteness. Consequences of a non-positive semi-definite covariance matrix are singular kriging systems and negative estimation errors.

Markov Models

Two models exist under this heading: Markov Model I and Markov Model II. The former is the more common Markov assumption used in most collocated cokriging applications, while the latter is a variation of the original model for cases where the volume support of the secondary data is much larger than that of the primary data.

Markov Model I. Modeling direct and cross-variograms is a complex and tedious task. A Markov-type model of coregionalization simplifies this process. Consider two standard jointly Gaussian RVs, $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u}), i \neq j$, which are the primary and secondary variable, respectively. The Markov-type assumption states that collocated hard data will screen the influence of other hard data that is further away [4, 81], i.e.

$$E\{Z_j(\mathbf{u}) | Z_i(\mathbf{u}) = z, Z_i(\mathbf{u}+\mathbf{h}) = z'\} = E\{Z_j(\mathbf{u}) | Z_i(\mathbf{u}) = z\} \tag{2.20}$$

Derivation of the Markov cross covariance model is based on determining the covariance of $Z_j(\mathbf{u})$ given $Z_i(\mathbf{u}) = z$ and $Z_i(\mathbf{u}+\mathbf{h}) = z'$, where $f_{\mathbf{h}}(z, z')$ is the bivariate pdf of $Z_i(\mathbf{u})$ and $Z_i(\mathbf{u}+\mathbf{h})$:

$$\begin{aligned} C_{ij}(\mathbf{h}) &= E\{Z_j(\mathbf{u}) \cdot Z_i(\mathbf{u}+\mathbf{h})\} \\ &= \int\int E\{Z_j(\mathbf{u}) \cdot Z_i(\mathbf{u}+\mathbf{h}) | Z_i(\mathbf{u}) = z, Z_i(\mathbf{u}+\mathbf{h}) = z'\} f_{\mathbf{h}}(z, z') dz dz' \\ &= \int\int z' E\{Z_j(\mathbf{u}) | Z_i(\mathbf{u}) = z, Z_i(\mathbf{u}+\mathbf{h}) = z'\} f_{\mathbf{h}}(z, z') dz dz' \\ &= \int\int z' E\{Z_j(\mathbf{u}) | Z_i(\mathbf{u}) = z) f_{\mathbf{h}}(z, z') dz dz' \text{ based on Markov assumption} \end{aligned}$$

Since the two RVs are jointly Gaussian, the regression of Z_j on Z_i is

$$E\{Z_j(\mathbf{u})|Z_i(\mathbf{u}) = z\} \quad = \quad \rho_{ij}(\mathbf{0}) \cdot z \ .$$

where $\rho_{ij}(\mathbf{0})$ is the correlation between $Z_i(\mathbf{u})$ and $Z_j(\mathbf{u})$ (i.e. collocated). This result gives the Markov cross covariance model:

$$\begin{aligned} C_{ij}(\mathbf{h}) &= \rho_{ij}(\mathbf{0}) \cdot \int\int z' z f_{\mathbf{h}}(z, z') dz dz' \\ &= \rho_{ij}(\mathbf{0}) \cdot C_{ii}(\mathbf{h}), \forall \mathbf{h} \end{aligned} \tag{2.21}$$

For standardized variables, that is, random variables with unit variance, Equation 2.21 becomes

$$\rho_{ij}(\mathbf{h}) = \rho_{ij}(\mathbf{0}) \cdot \rho_{ii}(\mathbf{h}), \forall \mathbf{h} \tag{2.22}$$

The Markov model only requires that the covariance function of the primary variable be modeled. The cross covariance between the primary and secondary variable is approximated using the relation in Equations 2.21 or 2.22. Use of only the collocated secondary data means that the covariance function of the secondary data is not required [4, 81].

One situation in which the Markov approximation is a poor assumption is the integration of data of significantly different volume supports. For example, suppose there is seismic and core data available at location **u**. The small scale data from the core sample cannot screen the seismic data that informs a much larger volume although both data are centered at the same location.

Markov Model II. For the case when the secondary variable $Z_j(\mathbf{u})$ is defined on a much larger volume support than the primary variable $Z_i(\mathbf{u})$, Journel introduced a variation of the Markov assumption referred to as Markov Model II [37]. Simply stated, Markov Model II assumes that collocated *secondary* data will screen the influence of other *secondary* data that is further away [37], i.e.

$$E\{Z_i(\mathbf{u})|Z_j(\mathbf{u}) = z, Z_j(\mathbf{u}+\mathbf{h}) = z'\} = E\{Z_i(\mathbf{u})|Z_j(\mathbf{u}) = z\} \tag{2.23}$$

The corresponding cross covariance model is

$$C_{ij}(\mathbf{h}) = C_{ij}(\mathbf{0}) \cdot C_{jj}(\mathbf{h}), \forall \mathbf{h} \tag{2.24}$$

Derivation of 2.24 follows from the derivation of Markov Model I. Unlike the more popular Markov Model I which requires only modeling of the covariance model of the primary variable, $C_{ii}(\mathbf{h})$ to define the cross covariance, Markov Model II requires that the covariance model of the secondary variable, $C_{jj}(\mathbf{h})$ be modeled.

The resulting cross-covariance is obtained via relation 2.24. Further, the Markov Model II defines the primary covariance as a function of the secondary covariance and a residual covariance, $C_R(\mathbf{h})$ [37, 65]:

$$C_{ii}(\mathbf{h}) = C_{ij}^2(0) \cdot C_{jj}(\mathbf{h}) + (1 - C_{ij}^2(0)) \cdot C_R(\mathbf{h}), \forall \mathbf{h} \tag{2.25}$$

Since Markov Model II requires modeling of the secondary covariance, from which subsequent definition of the primary and cross covariance is possible, the resulting model of coregionalization must be checked to ensure the model is positive semi-definite.

Markov-Bayes

Use of the LMC and the Markov models of coregionalization in traditional Gaussian algorithms only allows for transfer of linear, homoscedastic correlation. The Markov-Bayes model aims to account for the entire conditional distribution at each location $\mathbf{u}$ [42, 83]. This model was developed for the purpose of improving non-parametric geostatistics, specifically indicator simulation.

Suppose Z_i is the sparsely sampled primary variable, and Z_j is the densely sampled secondary variable. The primary data are considered "hard" data and are coded as indicators:

$$i(\mathbf{u}_\alpha; z) = \begin{cases} 1, & \text{if } z_i(\mathbf{u}_\alpha) \leq z \\ 0, & \text{otherwise} \end{cases}$$

where $z_i(\mathbf{u}_\alpha)$ is the primary data value at location $\mathbf{u}_\alpha$. By convention the indicator random variable is denoted $I(\mathbf{u}_\alpha; z)$ with outcome $i(\mathbf{u}_\alpha; z)$, which is not to be confused with the variable subscript index "i" for the primary variable.

Secondary data $Z_j(\mathbf{u}_\alpha)$ are used to define a "local prior" or a "pre-posterior" distribution of $Z_i(\mathbf{u}_\alpha)$. Secondary data are coded as probabilities or Y data:

$$y(\mathbf{u}_\alpha; z) = prob\{Z_i(\mathbf{u}_\alpha) \leq z | Z_j(\mathbf{u}_\alpha)\}$$

where $y(\mathbf{u}_\alpha; z) \in [0, 1]$ and usually $\neq F_i(z)$. For locations where hard data exists (i.e. $z_i(\mathbf{u}_\alpha)$ is known), the local prior cdf becomes

$$y(\mathbf{u}_\alpha; z) = \begin{cases} 1, & \text{for all } z \leq z_i(\mathbf{u}_\alpha) \\ 0, & \text{for all } z > z_i(\mathbf{u}_\alpha) \end{cases}$$

Secondary information can also be coded as a constraint interval or a continuous interval based on calibration to a bivariate distribution representing correlation between primary and secondary data [42, 83].

This model requires (1) a Markov-type assumption to simplify modeling of cross covariance models between $I(\mathbf{u}; z)$ and $Y(\mathbf{u}; z)$, and (2) use of Bayes theorem to update local prior distributions and obtain posterior conditional distributions, given that direct and cross covariances are known.

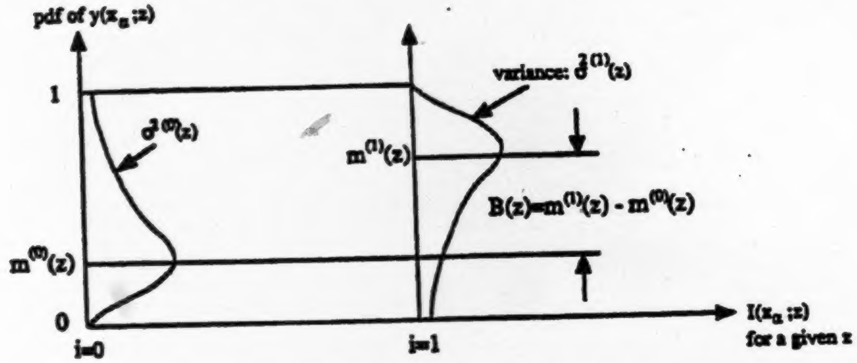


Figure 2.4: Graphical interpretation of calibration parameter $B(z)$ of soft data in Markov-Bayes updating. Source: Journel and Zhu, 1990

Based on the Markov approximation, the direct and cross covariances are calibrated by $B(z)$ (see Figure 2.4) [42, 83]:

$$C_{IY}(\mathbf{h};z) = B(z)\cdot C_I(\mathbf{h};z)\ \forall\mathbf{h}$$

$$C_Y(\mathbf{h};z) = \begin{cases} B^2(z)\cdot C_I(\mathbf{h};z), & \forall\mathbf{h}>0 \\ |B(z)|\cdot C_I(\mathbf{h};z), & \mathbf{h}=0 \end{cases}$$

where

$$\begin{aligned} B(z) &= m^1(z)-m^0(z) \\ m^1(z) &= E\{Y(\mathbf{u};z)|I(\mathbf{u};z)=1\} \\ m^0(z) &= E\{Y(\mathbf{u};z)|I(\mathbf{u};z)=0\} \end{aligned}$$

Zhu and Journel (1990, 1993) interpret the difference between $m^1(z)$ and $m^0(z)$ as a measure of accuracy of the local prior distributions of $Y(\mathbf{u};z)$ in predicting $Z_i(\mathbf{u})\le z$ and $Z_i(\mathbf{u})>z$, respectively [42, 83]. A calibration of $B(z)=1$ is considered the best in terms of accuracy since this means that the primary and secondary RV are perfectly spatially correlated, i.e. $C_Y(\mathbf{h};z)=C_I(\mathbf{h};z)$ and $C_{IY}(\mathbf{h};z)=C_I(\mathbf{h};z)$, $\forall\mathbf{h}$. Conversely, $B(z)=-1$ is interpreted as perfect "error" where the event $Z_i(\mathbf{u};z)\le z$ is actually assigned the probability of $Z_i(\mathbf{u};z)>z$. The worst case occurs when $B(z)=0$ which indicates that soft information $Y(\mathbf{u};z)$ does not help in predicting the value of the indicator $I(\mathbf{u};z)$.

Once the set of calibration parameters $B(z),\forall z$ is established, the model of coregionalization is fully defined. The Markov approximation along with Bayesian updating requires only the direct covariance of the primary variable or hard data

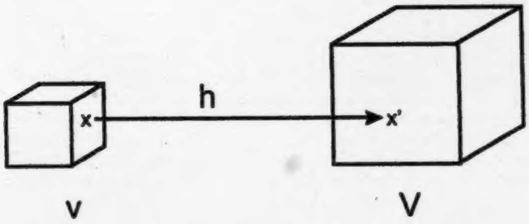


Figure 2.5: Schematic illustration for volume variance showing two volumes, v and V separated by a lag vector $\mathbf{h}$.

be modeled. As this model is dependent on a Markov assumption, it is also a poor approximation when conditioning to data of significantly different supports.

2.1.5 Volume Variance

Thus far, no mention has been made about the volume support of the data. Consider in a petroleum context, where well log and seismic data are available. Well log samples can be vertically spaced as close as 10cm apart, while seismic data can be found at 50-100m intervals. The log information is considered to inform a much smaller volume of the domain than the seismic data. To integrate different types of data, the support of the samples must be considered.

The average covariance ($\bar{C}$) is the mean value of the covariance calculated when one extremity of the lag vector $\mathbf{h}$ informs a volume v and the other extremity informs a larger volume V (see Figure 2.5).

$$\overline{C(V,v)} = \frac{1}{V \cdot v} \int_{V(\mathbf{u})} dx' \int_{v(\mathbf{u}')} C(x - x')dx \qquad (2.26)$$

Similarly, the mean variogram value is calculated as

$$\overline{\gamma(V,v)} = \frac{1}{V \cdot v} \int_{V(\mathbf{u})} dx' \int_{v(\mathbf{u}')} \gamma(x - x')dx \qquad (2.27)$$

The mean covariance and the mean variogram depend on the geometry of the two volumes, v and V, and the covariance/variogram functions. Analytical expressions can be obtained by solving the quadruple or sextuple integrals (for 2D and 3D, respectively) in Equations 2.26 or 2.27 for simple models like the spherical [13, 52] (see Figure 2.3). As well, some tables are available that summarize the values for the spherical and exponential variogram in 2D and 3D [40]. In practice, however, the common approach to calculating these two functions is numerical integration. This numerical integration is done by discretizing the two volumes into a number of points, calculating the variogram values between all combinations of the discretized points in v with the points in V, and then averaging these to obtain the mean value.

A large value of $\overline{\gamma(V,v)}$ indicates large variability between the volumes v and V. Note further that $\overline{\gamma(v,v)}$ is the variability between volume v with itself, and this value will decrease as the volume decreases. Note that if v is a point support, $\overline{\gamma(v,v)} = 0$ and $\overline{C(v,v)} = \sigma^2$. Conversely, if V is an infinite stationary domain, $\overline{\gamma(V,V)} = \sigma^2$ and $\overline{C(V,V)} = 0$.

A related notion to the expected value of the covariance and the variogram, is the dispersion variance, which is the expected value of the variance.

$$D^2(v,V) = E\left\{\left(\underbrace{z_i}_{\text{Support v}} - \underbrace{m_i}_{\text{Support V}}\right)^2\right\} \tag{2.28}$$

The dispersion variance is related to the mean covariance by

$$D^2(v,V) = \overline{C}\,(v,v) - \overline{C}\,(V,V) \tag{2.29}$$

and to the mean variogram by

$$D^2(v,V) = \overline{\gamma}\,(V,V) - \overline{\gamma}\,(v,v) \tag{2.30}$$

Since the covariance model is required to be positive semi-definite, $D^2(v,V) \geq 0$. If v is a point support, then the dispersion variance simplifies to $D^2(v,V) = \overline{\gamma}\,(V,V)$. Conversely, if V is an infinite domain, then $D^2(v,V) = \overline{C}\,(v,v)$.

An important property of the dispersion variance is the additivity of variances:

$$D^2(v,A) = D^2(v,V) + D^2(V,A) \tag{2.31}$$

This additivity property is also referred to as "Krige's relation". The variance of a small volume, v, within the domain A, is given by the sum of the variance of the small support within the block volume, V, and the variance of the block support within the domain. This relation quantifies the change in the expected variance with volume.

In practice, it is possible to use $\bar{C}$ directly to solve the kriging equations to account for differences between the support of the data and the volume to be estimated. Krige's relation is used to determine the change in the variance as the support changes.

2.1.6 Kriging

Kriging is an optimization technique consisting of a class of linear regression algorithms used in spatial estimation. In estimating unsampled locations, weights are assigned to the known dataset and the estimate is a linear combination of the sample data values.

This class of algorithms consists of many different "flavours" of kriging: simple, ordinary, block, cokriging, disjunctive [53], universal [31], multiGaussian [72, 73, 74], etc. As its name suggests, the most basic form of kriging is simple kriging (SK), wherein there are no constraints on the assigned values of the weights. Another commonly used form of kriging is ordinary kriging (OK), a technique in which the weights are constrained to sum to 1. Similar to OK, the other techniques are variations of the SK method and may account for potential trends in the data, volume support of the data, etc. The kriging equations are derived below for SK; OK is also addressed for its robustness in practice. The context of estimation using multiple variables is also considered.

In general, the kriged estimate is defined as

$$Z^*(\mathbf{u}) - \mu(\mathbf{u}) \quad = \quad \sum_{\alpha=1}^{n} \lambda_\alpha [Z(\mathbf{u}_\alpha) - \mu(\mathbf{u}_\alpha)] \tag{2.32}$$

where $Z^*(\mathbf{u})$ is the estimate at the location of interest, $\mathbf{u}$, λ_α is the weight given to the data value at location $\mathbf{u}_\alpha, \alpha = 1, \ldots, n$, and $\mu(\mathbf{u})$ and $\mu(\mathbf{u}_\alpha)$ denotes the expected values at locations $\mathbf{u}$ and $\mathbf{u}_\alpha$, respectively. In practice, n may vary from location to location, so that only a maximum number of nearby data are considered for estimation.

The weights are determined by minimizing the expected squared error between the true value, $Z(\mathbf{u})$, and the estimate, $Z^*(\mathbf{u})$:

$$\sigma_E^2 \quad = \quad E\{[Z(\mathbf{u}) - Z^*(\mathbf{u})]^2\} \tag{2.33}$$

under the condition of unbiasedness, that is,

$$E\{Z(\mathbf{u}) - Z^*(\mathbf{u})\} \quad = \quad 0 \tag{2.34}$$

Equations 2.32 to 2.34 form the basic equations for kriging. Interestingly, minimization of the error variance in Equation 2.33 is not dependent on the data values, but rather it is dependent on their spatial configuration and their covariance structure.

Simple Kriging

In SK, the mean is considered known and stationary. Hence the kriging estimate in Equation 2.32 can be written in terms of the residual variable, Y,

$$Y(\mathbf{u}_\alpha) \quad = \quad Z(\mathbf{u}_\alpha) - \mu(\mathbf{u}_\alpha), \qquad \alpha = 1, \ldots, n$$

to give the estimate

$$Y^*(\mathbf{u}) \quad = \quad \sum_{\alpha=1}^{n} \lambda_\alpha Y(\mathbf{u}_\alpha)$$

The kriging variance becomes

$$\begin{aligned}\sigma^2_{SK} &= E\{[Y(\mathbf{u}) - Y^*(\mathbf{u})]^2\} \\ &= E\{Y(\mathbf{u}), Y(\mathbf{u})\} - 2E\{Y(\mathbf{u}), Y^*(\mathbf{u})\} + E\{Y^*(\mathbf{u}), Y^*(\mathbf{u})\} \\ &= C(0) - 2\sum_{\alpha=1}^{n} \lambda_\alpha C(\mathbf{u}, \mathbf{u}_\alpha) + \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta C(\mathbf{u}_\alpha, \mathbf{u}_\beta) \qquad (2.35)\end{aligned}$$

Minimizing Equation 2.35 with respect to the weights yields the following SK system of equations:

$$\sum_{\beta=1}^{n} \lambda_\beta C(\mathbf{u}_\alpha, \mathbf{u}_\beta) = C(\mathbf{u}, \mathbf{u}_\alpha) \quad \alpha = 1, \ldots, n$$

The adopted model of coregionalization provides all the required information in order to solve the kriging equations. As a result, these weights account for closeness of the data to the point of estimation and the redundancy between data locations.

By construction, the kriging weights yield the minimum error variance between the true data, $Y(\mathbf{u})$ and the estimate, $Y^*(\mathbf{u})$,

$$\sigma^2_{SK} = C(0) - \sum_{\alpha=1}^{n} \lambda_\alpha C(\mathbf{u}, \mathbf{u}_\alpha)$$

Ordinary Kriging

Within the region of interest, the mean value of an attribute may vary locally and may be unknown. OK accounts for this by limiting the region of stationarity to within the neighbourhood centered at the location of interest [28], so

$$\mu(\mathbf{u}) = \mu(\mathbf{u}_\alpha), \quad \forall \alpha = 1, \ldots, n$$

Recall that SK requires that the mean is known and stationary, no such assumptions are made in OK, hence the kriging estimate in Equation 2.32 becomes

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha Z(\mathbf{u}_\alpha) + (1 - \sum_{\alpha=1}^{n} \lambda_\alpha)\mu(\mathbf{u}) \qquad (2.36)$$

The unknown mean is filtered by applying the following constraint

$$\sum_{\alpha=1}^{n} \lambda_\alpha = 1 \qquad (2.37)$$

which simplifies the OK estimate to

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_\alpha Z(\mathbf{u}_\alpha) \tag{2.38}$$

The kriging system of equations is then obtained by minimizing the estimation variance, subject to the constraint in Equation 2.37. In order to solve this, a Lagrangian parameter of 2μ is used and the Lagrangian function, $Q(\mathbf{u})$, becomes:

$$Q(\mathbf{u}) = \sigma_E^2 + 2\mu\left(\sum_{\alpha=1}^{n} \lambda_\alpha - 1\right) \tag{2.39}$$

The optimal weights are obtained by minimizing $Q(\mathbf{u})$, which requires that the partial derivatives be taken with respect to the weights and the Lagrangian parameter:

$$\begin{aligned} \frac{\partial Q(\mathbf{u})}{\partial \lambda_\alpha} = 0 &= \sum_{\alpha=1}^{n} \lambda_\alpha C(\mathbf{u}_\alpha, \mathbf{u}_\beta) - C(\mathbf{u}, \mathbf{u}_\alpha) + \mu \\ \frac{\partial Q(\mathbf{u})}{\partial \mu} = 0 &= \sum_{\alpha=1}^{n} \lambda_\alpha - 1 \end{aligned}$$

The OK system of equations then becomes

$$\begin{aligned} \sum_{\beta=1}^{n} \lambda_\beta C(\mathbf{u}_\alpha, \mathbf{u}_\beta) + \mu &= C(\mathbf{u}, \mathbf{u}_\alpha) \qquad \forall \alpha = 1, \ldots, n \\ \sum_{\alpha=1}^{n} \lambda_\alpha = 1 \end{aligned}$$

This results in a system of $n+1$ equations with $n+1$ unknowns. Solving this system of equations yields the optimal weights under the imposed constraint.

Kriging for Multiple Variables

For multiple data types, the formalism is referred to as cokriging. The following development is consistent with simple cokriging.

Consider estimating the primary variable $Y_i^*(\mathbf{u})$ using P data types (where i can be any one of the P variables):

$$Y_i^*(\mathbf{u}) = \sum_{p=1}^{P} \sum_{\alpha=1}^{n_p} \lambda_{\alpha p} Y_p(\mathbf{u}_{\alpha p})$$

The cokriging system of equations consists of $\sum_{p=1}^{P} n_p$ equations:

$$\sum_{q=1}^{P}\sum_{\beta=1}^{n_q} \lambda_{\beta q} C(\mathbf{u}_{\alpha p}, \mathbf{u}_{\beta q}) = C_{ip}(\mathbf{u}, \mathbf{u}_{\alpha p}), \qquad \forall p = 1, \ldots, P; \alpha = 1, \ldots, n_p$$

The corresponding estimation variance is

$$\sigma_E^2 = C_{ii}(0) - \sum_{p=1}^{P}\sum_{\alpha=1}^{n_p} \lambda_{\alpha p} C_{ip}(\mathbf{u}, \mathbf{u}_{\alpha p})$$

In the general case of multiple variables at multiple volume scales, consideration must be given to the generalized cokriging equations. A thorough treatment of this system of equations along with a numerical example is provided in Appendix B.

Remarks on Kriging

One of the characteristic properties of kriging is that it is an exact interpolator. An estimate at a sample data location will result in the exact value of the sample data; the weights assigned to the other data is zero, and a weight of 1 is assigned to the data at the estimate location. A second notable property is that it is an unbiased estimator (Equation 2.34). Kriging is often referred to as the best linear unbiased estimate (BLUE); it is considered "best" in the sense of minimization of the expected squared error.

There is however an undesirable effect produced from kriging. Estimating an intermediate point between two high values will lead to a high estimate and show no indication of the potentially low-valued region in between. Kriging produces maps that are smoother than the reality, thus introducing a higher degree of spatial correlation than actually exists.

2.1.7 Simulation

The literature in this area is extensive. For numerical modeling of natural phenomena, there are numerous simulation algorithms that are available. These techniques include Gaussian, indicators, p-field, direct, simulated annealing, etc. By far, the most common and simplest methods are the Gaussian-related simulation algorithms. The main part of this section will focus on the different approaches to Gaussian simulation, and a brief discussion is provided on a few of the other simulation methods.

In classical geostatistics, simulation requires the definition of a conditional cumulative distribution function(ccdf) and Monte Carlo simulation (MCS) from this ccdf. The kriging estimate and variance are used to estimate local ccdfs. The main difference between sequential simulation and kriging is that simulation uses both simulated locations and data locations to determine the kriging weights. This allows the covariance between the simulated values to be correct [17]. Monte Carlo simulation is straightforward; the discussion in this section will be limited to different types of simulation approaches.

Gaussian Simulation

Gaussian simulation is one of the simplest geostatistical simulation algorithms, and for this reason, it is the most commonly used method in practice. This technique is dependent on the characteristics of a Gaussian or normally distributed variable and an assumption of multiGaussianity. This assumes that the multivariate distribution is also Gaussian, thus it must follow a very particular functional form:

$$f(Z) \quad = \quad \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \cdot exp\left[-(Z-\mu)'\Sigma^{-1}(Z-\mu)/2\right] \tag{2.40}$$

where Z is a random vector of n random variables $[Z_1, \ldots, Z_n]'$, Σ is the $n \times n$ covariance matrix, and μ is a $1 \times n$ matrix of the means of each random variable Z_i [35].

The advantage of the multivariate Gaussian (or normal) distribution is that all conditional and marginal distributions are also Gaussian, which are fully defined by knowing only the mean and variance. This is what makes Gaussian simulation so simple and popular in practice. Monte Carlo simulation of the resulting ccdf provides the simulated value at a particular location. Simulation honours the local data, reproduces the histogram and spatial continuity, and allows for uncertainty assessment [17, 28, 40].

Important characteristics of the multivariate Gaussian distribution are linearity and homoscedasticity. As with most other geostatistical simulation algorithms, the assumptions of stationarity and ergodicity are applicable.

There are four common implementations of Gaussian simulation: moving average, turning bands, matrix and sequential. The sequential approach is widely applied in practice. A brief description of each approach is provided below. In all cases, data transformation prior to and after the simulation is required to ensure the input data are indeed Gaussian variables (Section 2.1.2).

Moving Average. The premise for this approach is to simulate the RF as a moving average [49]. If the covariance of a RF can be expressed as a convolution product of a weight function, f, with its transpose, then the RF can be simulated using this technique [7, 15]. In practice, it may be difficult to determine f. This technique requires (1) a second order stationary RF, $X(\mathbf{u})$ with known covariance $C_x(\mathbf{h})$, (2) use of a Fourier transform, F, such that $C_y(\mathbf{h}) = C_x(\mathbf{h}) * F$, where $C_y(\mathbf{h})$ is the covariance of the second order stationary RF of interest, $Y(\mathbf{u})$.

The general procedure to perform a conditional simulation:

- Define a second order stationary RF $X(\mathbf{u})$ with $E\{X(\mathbf{u})\} = 0$ and corresponding $C_x(\mathbf{h})$.
- Define RF $Y(\mathbf{u})$ such that $Y(\mathbf{u})$ is a weighted average of $X(\mathbf{u})$.

$$Y(\mathbf{u}) = \sum_{\alpha=1}^{n} f(\mathbf{u}_\alpha)x(\mathbf{u}_\alpha)$$

Luster shows the above expression yields the covariance of $Y(\mathbf{u})$, $C_y(\mathbf{h})$ [49].

- Perform deconvolution on F to get weighting function f.

$$F = C_y(\mathbf{h})C_x(\mathbf{h})^{-1}$$

- Go to each node, and calculate the unconditional simulated value, $z_{uc}(\mathbf{u})$, by applying the weighting function to the surrounding data.

A special case of this method occurs when $X(\mathbf{u})$ is a uniform random variable (with $C_x(\mathbf{h})$ defined by nugget model), and the variogram of $Y(\mathbf{u})$ is a spherical model with range a, the weight function f is found to be constant. This amounts to the case where all nodes within a diameter equal to the range is given equal weight. The simulated node is then calculated as the equal weighted average of all $X(\mathbf{u})$ found within the range.

The resulting simulated values may require a post-processing rescaling to ensure a standardized normal distribution (with zero mean and unit variance). Note further that this process yields an *un*conditional simulated value, which can be conditioned via the following equation

$$z_s(\mathbf{u}) \quad = \quad z_k^*(\mathbf{u}) + [z_{uc}(\mathbf{u}) - z_{uc}^*(\mathbf{u})] \tag{2.41}$$

This second step of conditioning the simulation essentially requires that kriging be performed twice: (1) using the data values, to get $z_k^*(\mathbf{u})$, and (2) using the unconditional simulated values at the same location as the data to get $z_{uc}^*(\mathbf{u})$. The simulated value, $z_s(\mathbf{u})$, at each node is calculated as in Equation 2.41, where $z_{uc}(\mathbf{u})$ is the unconditional simulated value obtained above (in general procedure).

Using this simple conditioning equation, notice that at the data location, $z_{uc}(\mathbf{u}) = z_{uc}^*(\mathbf{u})$ and the simulated value is simply the data value - as it should be. Beyond the range of correlation, $z_k^*(\mathbf{u})$ and $z_{uc}^*(\mathbf{u})$ will equal the global mean, and the simulated value will equal the unconditional simulated value, $z_{uc}(\mathbf{u})$.

Moving average methods are infrequently used due to high CPU requirements, especially if the number of nodes to be averaged is large. This is particularly true for the variogram models that reach the sill asymptotically, such as the exponential and the Gaussian variograms.

Turning Bands. In this implementation, the simulation space is populated with N lines uniformly distributed in a unit sphere. For a 3-D simulation, each of the N lines is assigned a 1-D realization based on the 1-D covariance of the variable [40]. The 1-D covariance is calculated by taking the partial derivatives of the 3-D covariance, $C(s)$:

$$C^1(s) = \frac{\partial}{\partial s} sC(s)$$

where $s = |\mathbf{h}|$ in one dimension [40, 49].

This 1-D realization can be generated using the moving average technique. Each line contributes equally to the total simulated value. The simulated value is then calculated as the sum of the N line contributions divided by $\sqrt{N}$ (to standardize the variance) (Equation 2.42).

It follows from this that the more lines used in the simulation, the more complex the simulation, but the better the RF covariance reproduction. Generally, 15 lines or more are used to define the 3-D space [40, 49]. Specifically, the procedure for this technique is:

- Generate N lines uniformly distributed in a sphere.
- For each line, generate a 1-D realization using the 1-D covariance of the variable.
- At each location, the contribution of each line is summed:

$$z_{uc}(\mathbf{u}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} z_i(\mathbf{u}) \tag{2.42}$$

This technique generates an unconditional realization with an isotropic variogram model. Simulating geometric anisotropy requires repeated application of the algorithm for a 1-D, 2-D and a 3-D isotropic variogram or covariance. The simulated value is then a sum of four simulated covariance models multiplied by its variance contribution (the 1-D, 2-D, 3-D isotropic variograms and a nugget model) [40].

Conditioning a turning bands simulation is done in the same manner as the moving average approach (see Equation 2.41).

Matrix Simulation. This approach is often referred to as LU simulation [3, 16] based on the fact that any matrix that is symmetric and positive definite can be decomposed via Cholesky decomposition [11], which results in a special case where $\mathbf{L}^T = \mathbf{U}$, where $\mathbf{L}$ is the lower triangular matrix, $\mathbf{U}$ is the upper triangular matrix, and the superscript T denotes the transpose of the matrix.

A large covariance matrix, $\mathbf{C}$ (consisting of the covariance between data to data, node to node and data to node), is defined that accounts for the entire simulation grid.

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}$$

where $\mathbf{C}_{11}$ is the $n \times n$ data to data covariance, $\mathbf{C}_{12}$ is the $n \times N$ data to grid node covariance, $\mathbf{C}_{21}$ is the $N \times n$ grid node to data covariance, and $\mathbf{C}_{22}$ is the $N \times N$ node to node covariance. Note that $\mathbf{C}_{12} = \mathbf{C}_{21}^T$. Since covariance matrices satisfy the requirement for symmetry and positive semi-definiteness, then it is possible to solve for its $\mathbf{L}$ or $\mathbf{U}$ matrix (getting either $\mathbf{L}$ or $\mathbf{U}$ will automatically give the other, since one is the transpose of the other).

$$\mathbf{C} = \mathbf{LU} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{0} & \mathbf{U}_{22} \end{bmatrix}$$

Once the laborious task of getting **L** is completed, unconditional simulation is simply a matter of matrix multiplication of the $\mathbf{L}_{22}$ with an $N \times 1$ column vector, ω, of independent normal deviates:

$$\mathbf{y} = \mathbf{L}_{22}\omega$$

Note that in the case of an unconditional simulation, the covariance matrix, **C**, consists only of the node to node covariance matrix, $\mathbf{C}_{22}$. Checking the expected value between **y** and $\mathbf{y}^T$ shows that the covariance matrix between the simulated locations, **C**, is reproduced.

$$\begin{aligned} E\{\mathbf{y}\mathbf{y}^T\} &= E\{\mathbf{L}\omega\omega^T\mathbf{L}^T\} \\ &= \mathbf{L}E\{\omega\omega^T\}\mathbf{L}^T \end{aligned}$$

Since

$$E\{\omega_i\omega_j\} = \begin{cases} 1, & \text{if } i = j, \forall i, j = 1, \ldots, N \\ 0, & \text{otherwise} \end{cases} \tag{2.43}$$

Then

$$E\{\mathbf{y}\mathbf{y}^T\} = \mathbf{L}\mathbf{I}\mathbf{L}^T = \mathbf{L}\mathbf{U} = \mathbf{C} \tag{2.44}$$

Thus, drawing a different vector of random normal deviates produces a different realization. The realization can be conditioned using Equation 2.41, which is the same procedure as applied for the moving average and the turning bands algorithms. Alternatively, Davis proposed decomposing a larger covariance matrix to give the following matrix multiplication for generating a conditional simulation all in one step [16]:

$$\begin{aligned} \mathbf{y} &= \mathbf{L}\omega \\ \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} \end{aligned}$$

where $\mathbf{y}_1$ is an $n \times 1$ column vector of the normal scores of the conditioning data, $\mathbf{y}_2$ is an $N \times 1$ column vector of required simulated values, $\omega_1 = \mathbf{L}_{11}^{-1}\mathbf{y}_1$, and ω_2 is an $N \times 1$ vector of independent normal deviates. Substitution of $\mathbf{L}_{11}^{-1}\mathbf{y}_1$ for ω_1 into the above system yields

$$\mathbf{y}_2 = \mathbf{L}_{21}\mathbf{L}_{11}^{-1}\mathbf{y}_1 + \mathbf{L}_{22}\omega_2$$

As before, the generation of a different realization only involves drawing another $N \times 1$ vector of random normal deviates for ω_2. This approach is CPU intensive, but it is particularly efficient if the grid is relatively small (less than a few hundred nodes) and a large number of realizations is required. It does not use a search radius.

Sequential Approach. This simulation approach is common, and the simplest technique [32]. This approach amounts to the application of Bayes Theorem to decompose the multivariate distribution into a series of conditional distributions for each location.

In sequential simulation, all previously simulated nodes are added to the data for determination of the ccdf at the next location, and MCS is used to draw from the ccdf defined by kriging. The simulated value, $z_s(\mathbf{u})$, is a sum of the kriged estimate, $z^*_{sk}(\mathbf{u})$, and a residual value, $r_s(\mathbf{u})$, that is drawn from a N(0,$\sigma^2_{sk}(\mathbf{u})$) distribution where $\sigma^2_{sk}(\mathbf{u})$ is the kriging variance.

$$z_s(\mathbf{u}) = z^*_{sk}(\mathbf{u}) + r_s(\mathbf{u})$$

Essentially, the simulated value is drawn from a N($z^*_{sk}(\mathbf{u})$,$\sigma^2_{sk}(\mathbf{u})$) distribution.

For practical applications, computational time may be reduced if the number of data used in the estimation (including previously simulated nodes) is limited. The only potential drawback is that restricting the number of data in kriging may cause the covariance to be poorly reproduced at large distances. Use of a multiple grid simulation can mitigate this effect [71].

The actual simulation consists of the following steps (note the steps listed below do not include the required pre- and post-processing step of data transformation):

1. Determine a random path visiting each node in the grid.
2. At each node:
    (a) Using nearby data and any previously simulated grid nodes, perform kriging to construct a conditional distribution (normally distributed with mean and variance given by kriging).
    (b) Draw a simulated value from this conditional distribution.
3. Repeat Step 2 until all locations have been simulated.

Repeat this procedure for multiple realizations. Use of a random path avoids any artefacts in the simulated values as a result of the regular path chosen [21, 32, 54].

Multivariate Gaussian Simulation. There are several implementations of multivariate Gaussian simulation. Simultaneous simulation of multiple variables at multiple locations can be achieved using a large LU decomposition [3, 16]. This essentially is an extension of the LU decomposition for one variable, and involves decomposing a covariance matrix that accounts for multiple variables [28].

Simulation of multiple variables at multiple locations can also be performed in a sequential manner via the common sequential Gaussian simulation [32]. Typically, the first variable is simulated, and then the second variable is simulated using the first variable as secondary information.

The primary advantage of the sequential technique over LU simulation is computational effort. For a large grid, the full covariance matrix is large and Cholesky decomposition of this large matrix is computationally expensive. Moreover, numerical precision limits the matrix size to 5-10 thousand.

An important decision in multivariate modeling is the choice of the model of coregionalization (Section 2.1.4). Adoption of either the LMC or the simpler Markov model will affect the variography component of the model construction. The resulting variogram model(s) provide the required information to define the necessary covariance matrices.

Alternative Non-Gaussian Simulation Approaches

There are many other stochastic simulation techniques that exist in modern geostatistics. In particular, three non-Gaussian simulation techniques are briefly described in this section: indicator, p-field, and direct simulation.

Indicator Techniques. Outside of the Gaussian framework, indicator approaches are among the most commonly used methods [36]. The indicator transform for a continuous RV, Z is

$$I(\mathbf{u};z) = \begin{cases} 1, & \text{if} \quad Z(\mathbf{u}) \le z \\ 0, & \text{if} \quad Z(\mathbf{u}) > z \end{cases}$$

with mean and variance given by

$$E\{I(\mathbf{u};z)\} = p$$

$$\sigma^2 = p(1-p)$$

The indicator transform is determined at a number of different thresholds. A variogram model is fitted for each threshold, and is interpreted as the transition probability for that threshold. Sequential indicator simulation (SIS) follows the same steps as the sequential Gaussian approach above. The main difference lies in the determination of the local conditional distributions. In SIS, the ccdfs are determined by kriging each threshold to obtain an estimate of the transition probability. The probabilities from all thresholds forms the ccdf, from which a simulated value will be drawn via MCS.

Since kriging is performed for multiple thresholds at each location, the SIS algorithm is more computationally expensive than the Gaussian method which only requires solving only one kriging system to determine the ccdf. On the other hand, the SIS approach allows non-parametric determination the ccdf, without the requirement for any multiGaussian assumptions.

P-field Simulation. Probability field (p-field) simulation [25] dissociates the stochastic simulation process into two components: (1) determination of the local conditional distributions via geostatistical estimation techniques, and (2) generation of a stochastic uniform random field with a prescribed correlation structure, that is, the correlated p-field. At each grid location, a realization of the RV is drawn using the probability value and the known local cdf.

The main advantage of this simulation technique is speed [68]. The initial step of identifying the conditional distributions can be performed using kriging. This step only has to be performed once. The second step of generating the p-field is generated by unconditional simulation of the random field. Problems associated to the simulation include the occurrence of local extremas near conditioning data [60, 68], and increased spatial continuity at short scale distances [60].

Direct Sequential Simulation (DSS). The idea of direct simulation is to generate realizations without transforming the data, that is to simulate the data *directly*. The basic algorithm follows the sequential approach to simulation [80].

There are several reasons for diverging from the common Gaussian simulation. Direct simulation allows for direct accounting of multiscale data. Use of the kriging equations to determine the conditional distributions in conventional Gaussian simulation does not permit reproduction of non-stationary features, such as the proportional effect (wherein the variability of the variable depends on its magnitude) or heteroscedasticity. Use of a rescaled variance (to account for the local mean) in DSS should be able to account for the proportional effect [57].

The main obstacle in this approach is the inference of the shape of the conditional distribution. Knowing only the kriged estimate and variance does not provide sufficient information regarding the conditional distribution (unless the variables are Gaussian). Recent publications have focussed on this portion of the simulation process [8, 10, 66, 18, 59].

Most recently, Deutsch et. al. proposed the identification of conditional distributions using normal quantile transformations [18, 59]. The transform between the global histogram in original space and its normal transform is well understood. In Gaussian space, the conditional distributions associated to a certain mean and variance are easy to obtain. The idea then is to perform a reverse quantile transform of the conditional distribution from Gaussian space to find its counterpart in original data space. Using this approach, a database of conditional distributions (in original space) can be prepared. This database can then be used to simulate in direct space without the requirement for data transformation.

2.2 Multivariate Statistics

The preceding sections described some of the key concepts in multivariate geostatistical theory. Restrictive linear models have resulted in simplified approximations for practical purposes. As a result, the practice of multivariate geostatistics is a balancing act between simple implementation, computational time constraints and adherence to theoretical foundations.

There are many multivariate statistical tools that have been relatively untapped in geostatistical theory development and application. The following sections provide an overview of multivariate statistical techniques that have considerable potential for use in a geostatistical framework.

The first class of techniques are dimension reducing methods that includes principal components analysis (PCA) and factor analysis (FA). These seek to simplify the multivariate problem by reducing the dimension of the data. Another class of approaches is data transformation. In particular, alternating conditional expectation (ACE) is discussed. The next group of techniques are classification techniques that include both discriminant analysis and cluster analysis. These approaches focus on the classification of data into groups.

2.2.1 Dimension Reduction Methods

These techniques aim to reduce the dimension of the data, thus simplifying the multivariate problem. The two primary approaches are principal components analysis and factor analysis. The development of the former technique is credited to the work of Pearson in 1901 and Hotelling in 1933, while the concept of factor analysis first originated with work by Spearman in 1904 and 1926 [43].

Principal Components Analysis. The basis for principal components analysis (PCA) is the transformation of correlated variables into uncorrelated variables called principal components. The principal components, Y_j, $j = 1, \ldots, P$, are linear combinations of the original variables, Z_i, $i = 1, \ldots, P$.

$$Y_j = \sum_{j=1}^{P} a_{ij} Z_i, \forall j$$

Spectral decomposition of the covariance matrix of the original variables yields its eigenvectors and eigenvalues. The $P \times P$ matrix of coefficients $a_{ij}, i, j = 1, \ldots, P$ is called the transformation matrix and is the matrix of eigenvectors. Since the covariance matrix is a positive definite matrix, all the eigenvalues are positive and are interpreted as the variance of the principal components. The importance of a principal component is derived directly from the rank of the eigenvalue, that is, the largest eigenvalue corresponds to the principal component with maximum contribution to the variance of the original data and is referred to as the first principal component. If the variability of the data can be adequately captured by consideration of only the first few principal components, then the dimension of the original multivariate problem is reduced.

One consequence of calculating uncorrelated variables that maximize the variance is sensitivity to outliers. Outliers inflate the variance of the data; therefore, the principal components may not be representative of the true underlying variable.

Past geostatistical experience with PCA includes indicator kriging and simulation using principal components [69, 70], the study of spatial correlation of principal components [27], and general inclusion in multivariate geostatistics literature [76].

Factor Analysis. The goals of factor analysis (FA) are similar to those of PCA: to reduce the dimension of the data by finding uncorrelated variables. The uncorrelated variables are referred to as *factors* in this approach. Unlike PCA where the principal components are linear combinations of the data, FA defines each random variable

$Z_i, i = 1, \ldots, P$ as a linear combination of K underlying and unobservable factors (where $K < P$), $f_k, k = 1, \ldots, K$ which are common to all P variables, plus an independent "error" term, $\varepsilon_i, i = 1, \ldots, P$, specific to Z_i.

$$Z_i - \mu_i = \sum_{k=1}^{K} l_{ik} f_k + \varepsilon_i \qquad i = 1, \ldots, P$$

where l_{ik} is the factor loading on the k^{th} factor for the i^{th} variable.

This technique is dependent on several key assumptions: (1) the common factors, $f_k, k = 1, \ldots, K$, are assumed to be uncorrelated with zero mean and unit variance; and (2) the specific "error" term $\varepsilon_i, i = 1, \ldots, P$ has a mean of zero and is assumed to be independent of the common factors and each other. No assumption is made about the variance of the specific term. Arising from these assumptions is another fundamental difference between the two methods: factor analysis is based on a specific statistical model while PCA is not based on any statistical model [51].

Since both the common and specific factors are uncorrelated with each other, the covariance between the RVs Z_i are explained solely by the factor loadings, l_{ik}. Furthermore, if $Z_i, i = 1, \ldots, P$ are standardized random variables, the correlation between two variables Z_i and Z_j is given by:

$$r_{ij} = \sum_{k=1}^{K} l_{ik} l_{jk} \tag{2.45}$$

Thus, two variables are highly correlated if the factor loadings for both variables are high for the same factors [12, 51]. For this reason, factor analysis is associated with finding the factors that contribute to maximal *covariance* while PCA is concerned with finding components that contribute to maximal *variance*. In the case of $i = j$, Equation 2.45 quantifies the amount of the variance of Z_i that is accounted for by the K common factors; this sum of squares of the loadings on the K factors of Z_i is referred to as the *communality* on Z_i [35].

Common estimation methods such as the principal component method and maximum likelihood method are used to determine the initial common factors and the specific factors [35]. Once determined, analysis of the communalities and specific loadings may require that the factors be rotated to simplify interpretation of these new variables. Factor rotation finds new factors that describe the data equally well, and are linear combinations of the initial factors. Rotation may be orthogonal (producing uncorrelated factors) or oblique (resulting in correlated factors).

Remarks. From the previous sections, we note that PCA is a variance-oriented technique while FA is covariance-oriented. There exists a very specific statistical model within the FA approach, while PCA is not dependent on any particular model. The solution obtained from PCA is unique and exact; while FA produces several possible solutions, owing to the available options in factor extraction and rotation methods. This makes PCA a more attractive technique to most statisticians.

Furthermore, Seber (1984) notes that factor analysis is not suitable for categorical data [64] and attempts at non-linear FA have resulted in some theoretical and practical difficulties [43]. Due to the slight differences in the methods, there is considerable confusion between the techniques in the literature. Overall though, PCA is more commonly applied while factor analysis seems to remain popular in its founding discipline of psychology.

It is important to note that although PCA and FA produce uncorrelated variables, the variables are neither independent nor Gaussian. The place of these methods in geostatistics may be limited to reducing the number of variables to simulate. Undoubtedly, reducing the dimension of the problem will simplify inference of the coregionalization model, which is often the most cumbersome task in multivariate geostatistics.

2.2.2 Data Transformation Methods

This section focuses on the transformation of one set of variables to another set that simplifies both analysis and simulation. Although the techniques discussed in the previous section could also be considered data transformation techniques, the distinction is made in that this group of methods does not attempt to reduce the dimension of the problem.

There are many univariate transformation techniques, including the normal score transform (Section 2.1.2), power law transformation, orthogonal polynomials, etc.; however, multivariate transformation techniques are not so abundant. A robust non-parametric multivariate transform called alternating conditional expectation is discussed. Other methods are not presented here such as data re-expression [49, 50] and logratio transformation [2] for removal of constraints in multivariate distributions. The latter transform is commonly used to account for the constant sum constraint, which is typical of percentage data and ratios wherein all variables are constrained to sum to 100% or 1.0, respectively [2, 49, 50]. In the instance where all the constituent variables are not available to sum to 1.0, one idea may be to create a remainder variable to absorb the missing proportion. This is commonly referred to as a "remaining-space" type of transformation [1, 22, 49]. The stepwise conditional transformation for transforming non-Gaussian multivariate distributions to multiGaussian distributions will be developed later in Chapter 3 [46, 49, 62].

Alternating Conditional Expectation (ACE). The alternating conditional expectation (ACE) algorithm was first introduced by Brieman and Friedman (1985) [9], as a non-parametric transformation that requires no assumption about the functional form of the multivariate distribution.

We begin by defining RVs $Y, Z_1, \ldots, Z_P$, where Y is a response variable and $Z_1, \ldots, Z_P$ are the predictor variables. Arbitrary functions $\theta(Y), \phi_1(Z_1), \ldots, \phi_P(Z_P)$ with zero mean corresponding to these variables are also defined. The theoretical basis of the algorithm assumes the distributions for RVs $Y, Z_1, \ldots, Z_P$ are known, and $E\{\theta^2(Y)\} = 1$. Regression of $\theta(Y)$ is performed using $\sum_{i=1}^{P} \phi_i(Z_i)$. The fraction of the variance not explained by regression is quantified as:

$$e^2(\theta(Y), \phi_1(Z_1), \ldots, \phi_P(Z_P)) = \frac{E\left\{[\theta(Y) - \sum_{i=1}^{P} \phi_i(Z_i)]^2\right\}}{E\{\theta^2(Y)\}} \tag{2.46}$$

Optimal transformations are chosen as those that minimize Equation 2.46 with respect to *all* functions $\theta(Y), \phi_1(Z_1), \ldots, \phi_P(Z_P)$.

The ACE algorithm is an iterative procedure that is used to find the optimal transformations $\theta^*, \phi_1^*, \ldots, \phi_p^*$. In the bivariate case, the optimal transformations, θ^* and ϕ_1^*, minimize:

$$e^2(\theta(Y), \phi(Z)) = E\{[\theta(Y) - \phi(Z)]^2\} \tag{2.47}$$

which is equivalent to maximizing the correlation between Y and Z, $\rho^*(Y, Z)$. In the multivariate case, this method is aimed at finding the optimal transformations that make the relationship between $\theta(Y)$, $\phi_1(Z_1), \ldots, \phi_P(Z_P)$ as linear as possible. This facilitates the application of conventional geostatistical techniques (which assume a linear relationship between the model variables).

For illustrative purposes, 200 data values were generated using the model $y = exp(1+2x)+\varepsilon$, where x is U(0,2) and ε is N(0,10). Figure 2.6 shows the cross plots between the original data, the predictor variable and its transform, the response and its transform, and the transformed response vs. the transformed predictor. Clearly, the bivariate distribution of the transformed variables is more linear than that of the original variables. Non-linearity between the original variables has been removed without assuming the form of the data distribution.

Thus far, applications in geostatistical simulation are limited, although ACE has been applied in reservoir characterization to integrate seismic data [82], to estimate permeability from well logs [29], and to model surfaces [84].

2.2.3 Classification Techniques

Up to now all the techniques have been concerned with finding an alternate set of variables that can be used to simplify geostatistical modeling. In this section, we shift to methods that are useful in grouping data to form different populations. Identification of different populations may be useful in improving assumptions of stationarity.

Discriminant Analysis. Discriminant analysis involves two types of multivariate data: (1) a set of groups with known distribution, and (2) a set of data with no *a priori* information about the group to which it belongs [26]. The objective of discriminant analysis is to reconcile the second type of data with the first type based on the different observations on each sample. In a geological context, this technique may be useful to determine the properties that characterize different facies or rock types. Based on these properties, samples taken from unidentified facies can be classified. This could be used in subsequent geostatistical analysis.

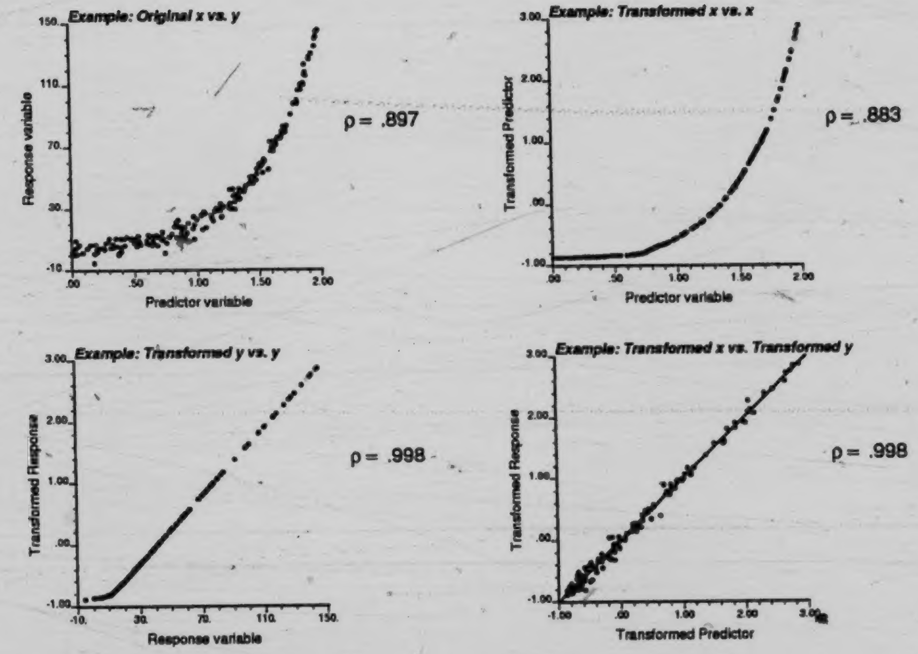


Figure 2.6: ACE Transformation of $y = exp(1 + 2x) + \varepsilon$, where x is U(0,2) and ε is N(0,10). Crossplots showing original y vs. x (top left), transformed x vs. x (top right), transformed y vs. y (bottom left), and transformed y vs. transformed x (bottom right).

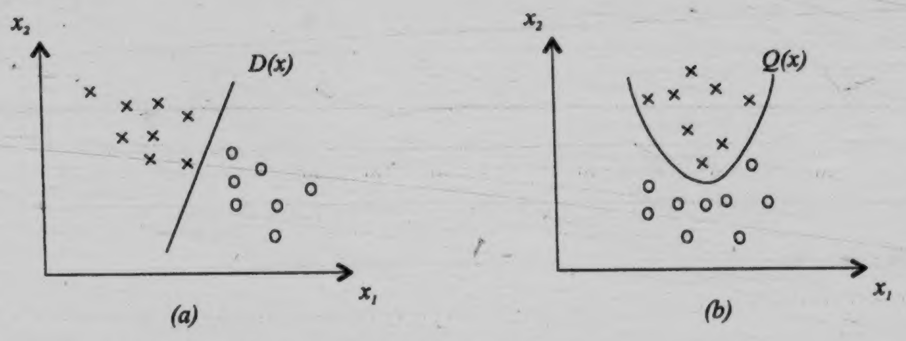


Figure 2.7: Use of (a) linear and (b) quadratic discriminant analysis to separate groups of data. Source: Seber (1984)

The first step in discriminant analysis is to represent the observations that clearly fall within the different $\pi_i, i = 1, \ldots, G$ groups. Once done, this set of data then becomes the measure by which the groups are characterized. For example, data known to be sampled within a certain facies, say sandstone, would belong to the sandstone group. Likewise, samples taken within shale would be grouped and identified as the shale facies group.

The second step is to determine the variables that best discriminates between the two or more groups. The focus is now shifted to finding the *right* measure to classify the unknown data. Many techniques exist that classify the data based on different criteria; these include linear discriminant analysis, classification by nearest neighbour methods, classification into one of several groups and classification using Mahalanobis distances. Seber (1984) provides a schematic illustration of the linear and quadratic discriminant analysis techniques (Figure 2.7) [64].

The cost of misclassifying the data is another important concept in discriminant analysis. The determination of the *right* classification rule is an optimization problem wherein the cost associated to classifying a sample to the wrong group must be minimized.

The suitability of classification rules depends on the data distributions, the cost associated to misclassification, and the goals of the classification rules (ranging from minimizing the number of samples misclassified to reducing the actual error rate for classifying future samples).

Discriminant analysis is concerned with assigning samples to *pre-established* groups [12]. Unfortunately, it will not help identify new groups. In practice, the benefits of applying discriminant analysis to geostatistical applications may be limited to classification of incomplete samples to known facies groups.

Cluster Analysis. At the start of any study, little may be known about the data or the population(s) from which they came. The main objective of cluster analysis is to identify groups within a set of data with n samples on which there are P-variate observations. The groupings identify samples that are similar based on the P-variate observations and distinguish them from those that are dissimilar. Each sample is

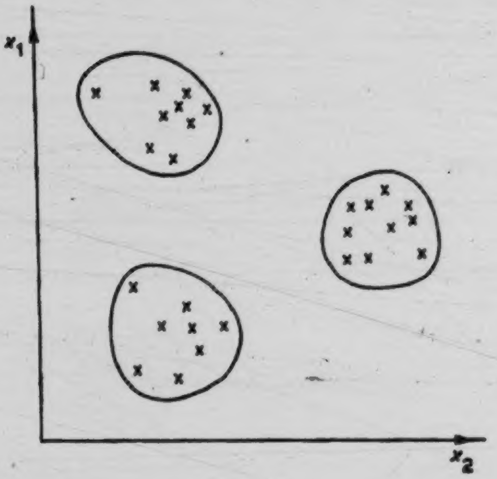


Figure 2.8: Clustering data into groups. Source: Chatfield & Collins (1980)

assigned to one group only and is considered dissimilar to samples belonging to other groups.

Cluster analysis is not limited to the separation of data samples, we could also cluster variables so that highly correlated variables are grouped together so that some combination of the variables could be used in analysis [12]. We have already seen this type of clustering in the dimension reducing methods of Section 2.2.1, and so the following discussion will focus on the clustering of samples.

Clustering of data samples can be achieved via agglomeration or division [51]. Consider that the analyst wishes to work with m clusters or groups for a total of n data samples, where $m < n$. The process of agglomeration requires that at the start of analysis, each sample forms its own group of one. Groups that are close to each other are then combined to form a single group. This is done until all the individual samples are placed into m groups. Alternatively, the process of division begins with only one group, to which all samples belong. Samples that are far apart are then divided, this continues until the m groups are formed and all samples have been accounted for. Chatfield & Collins (1980) provides schematic illustrations of clustering of samples based on two variables, x_1 and x_2(see Figure 2.8).

If we continued to perform cluster analysis within the first set of groups and identify sub-groups within the groups and so on, then the result is a hierarchical clustering scheme. This scheme essentially breaks down the primary groups into secondary groupings, until eventually each sample is its own group which essentially amounts to clustering by division. Figure 2.9 shows a schematic illustration of a hierarchical tree (commonly referred to as a dendrogram), also taken from Chatfield & Collins (1980).

The decision to combine or divide groups is based on distance measures between the data and the group centres, which may be dependent on the means, variances

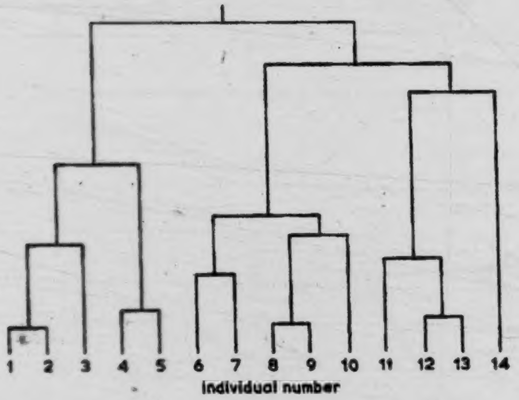


Figure 2.9: Hierarchical clustering: dendrogram or hierarchical tree. Source: Chatfield & Collins (1980)

and covariances of the two different populations. Metric distance measures are those which are strictly positive. Some common distance measures include Euclidean, Mahalanobis, Penrose, and Minkowski distance [23, 24, 35, 51, 67].

The main differences in the large number of cluster analysis algorithms lies in the choice of (1) the clustering process and (2) the distance measure applied.

Remarks. Cluster analysis may be useful in identification of rock types for rock type modeling. Discriminant analysis may be applied to classify incomplete samples to known rock types. Once rock types are identified and all samples are accounted for, modeling of continuous variables (e.g. grades or petrophysical attributes) can proceed via geostatistical simulation.

Both discriminant and cluster analysis should provide insight into the potentially different populations found within the data set. Decisions of stationarity may be applicable only to the individual groups identified using these techniques.

Chapter 3

Stepwise Conditional Transformation

The stepwise conditional transformation provides an alternative to the conventional normal score transform when multiple variables are considered for Gaussian simulation. This Chapter addresses the theoretical development of the technique, and explores the model of coregionalization associated to the new transformed variables.

The stepwise conditional transformation technique was first introduced by Rosenblatt in 1952 [62]. It is identical to the normal score transform in the univariate case. For bivariate problems, the normal scores transformation of the second variable is conditional to the probability class of the first variable. Correspondingly, for n-variate problems, the n^{th} variable is conditionally transformed based on the first $n-1$ variables, that is,

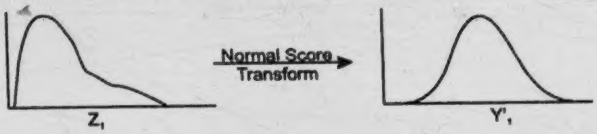
$$\begin{aligned} Y_1' &= G^{-1}[F_1(z_1)] \\ Y_2' &= G^{-1}[F_{2|1}(z_2|z_1)] \\ &\vdots \\ Y_n' &= G^{-1}[F_{n|1,\ldots,n-1}(z_n|z_1,\ldots,z_{n-1})] \end{aligned}$$

where $Z_i, i = 1,\ldots,n$ are the original variables and $Y_i', i = 1,\ldots,n$ are the corresponding stepwise conditionally transformed variables. Note that Y_i' refers to the stepwise conditional variables, while references to Y_i (that is, no superscript prime) denotes the conventional normal score transformed variables.

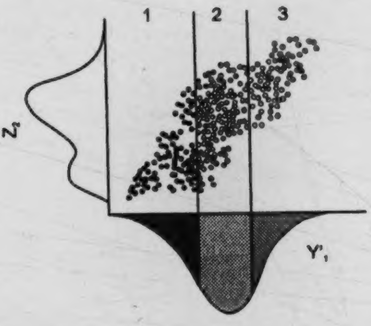
Figure 3.1 shows the steps to accomplish this conditional transformation for two variables. The primary variable is normal score transformed to yield Y_1'; note that for the primary variable *only*, Y_1' is the same as Y_1. The secondary data, Z_2, are binned into probability classes based on the paired primary data value, Y_1'. Each subset of secondary data is then normal score transformed. Since the grouping of secondary data is based on probability classes of the primary data, subsetting based on the original primary variable, Z_1, or the transformed primary variable, Y_1', will yield the same results. Transformation of three or more variables requires increased conditioning, but the procedure is just as straightforward as the bivariate case.

The result of this transformation is zero correlation between the transformed variables at $\mathbf{h} = 0$.

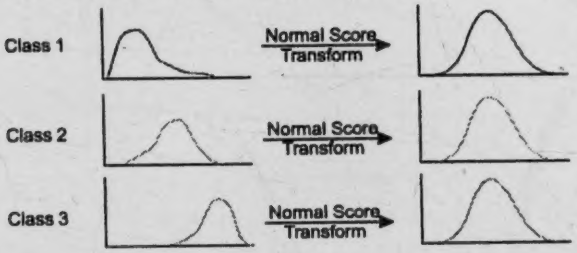
(a) Normal score transform Z_1 to yield Y'_1.


Normal Score
Transform
Z₁
Y'₁

(b) Partition Z_2 data into classes conditional to Y'_1



(c) Normal score transform each class of Z_2


Class 1
Normal Score
Transform
Class 2
Normal Score
Transform
Class 3
Normal Score
Transform

(d) Crossplot of stepwise conditionally transformed variables, Y'_1 and Y'_2

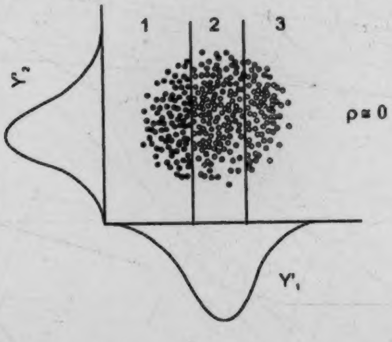


Figure 3.1: Schematic illustration of stepwise conditional transformation of two variables, Z_1 and Z_2, with Z_1 as the primary variable: (a) normal score transform the primary variable, Z_1; (b) partition Z_2 data based on classes of Y_1'; (c) perform normal score transform of each class of Z_2; (d) crossplot stepwise conditional variables to show bivariate Gaussian distribution with approximately zero correlation.

$$C(Y_i'(\mathbf{u}), Y_j'(\mathbf{u})) = C_{ij}'(0) = 0, \text{ for } i \neq j, i = 1, \ldots, n; j = 1, \ldots, n$$

Since each class of data is independently transformed to a normal distribution, correlation between Y_2' and Y_1' is removed. The marginal distribution of each transformed variable is Gaussian by construction. Moreover, all multivariate distributions are Gaussian in shape at distance lag $\mathbf{h} = 0$. This combination of zero correlation and multivariate Gaussianity are sufficient conditions for independence of the stepwise conditional variables (SC scores). Consequently, the simulation of a multivariate problem may not require cosimulation due to independence of the transformed variables.

By conditionally transforming the data, new variables are created that have no straightforward physical interpretation. The multivariate spatial relationship of the original model variable is not transformed (that is, no modification of bivariate spatial distributions $Z(\mathbf{u})$ and $Z(\mathbf{u}+\mathbf{h})$, or trivariate distributions $Z(\mathbf{u})$, $Z(\mathbf{u}+\mathbf{h}_1)$ and $Z(\mathbf{u}+\mathbf{h}_2)$, etc.).

The original multivariate distribution is honoured by back transformation via a reverse quantile transform. Back transformation of the n^{th} variable is conditional to the values of the first $n-1$ variables. For example, Z_1 can be determined from Y_1' with the correct conditional distribution; from the back transformed Z_1 and the simulated value of Y_2', Z_2 can be calculated; and so forth.

Figure 3.2 shows two mining examples for oil sands data and nickel laterite data. For each sample dataset, cross plots are shown for the original data, conventional normal score transformation, and stepwise conditional transformation. For the oil sands data, the cross plot of the normal scores shows an almost linear bivariate distribution with negative correlation. Application of the stepwise conditional transform yields a bivariate Gaussian distribution with almost no correlation. Conventional normal scores transformation of the nickel laterite data shows a positively correlated bivariate distribution that appears slightly heteroscedastic; while the stepwise conditional scores show a bivariate Gaussian distribution with essentially zero correlation.

Figure 3.3 shows two petroleum related examples which are referred to as the "Two Well" data and the "East Texas" data. The bivariate distributions after a direct normal scores transform are clearly more problematic than those obtained using the mining data. The normal scores cross plot on the left is heteroscedastic, while the cross plot on the right is non-linear and constrained in some fashion. After applying the stepwise conditional transformation, the bivariate distributions again exhibit a bivariate Gaussian distribution with essentially zero correlation.

Note that after stepwise conditional transformation, there are slight deviations from perfect bivariate Gaussian distributions in both Figures 3.2 and 3.3. This is particularly evident in the oil sands and the Two Well example. For the oil sands, there is evidence of spatial structure in the low values of the transformed bitumen ($\%bitumen < -1.28$); for the Two Well data, structure is apparent for the transformed porosity at high values ($\phi \geq 1.28$). These remnant structures reflect the structures within the classes (first and last class for the oil sands and the Two Well

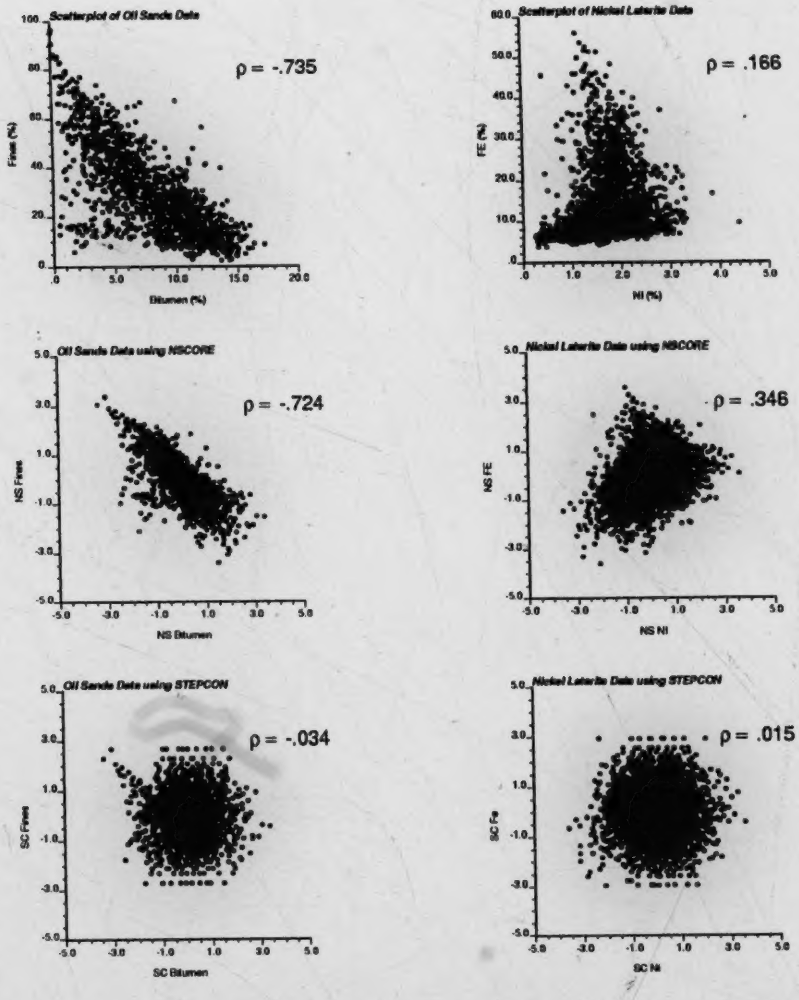


Figure 3.2: Comparative illustration of cross plots of the original data (top), normally transformed data (middle) and the stepwise conditionally transformed data (bottom) for Oil Sands data (left side) and Nickel Laterite data (right side).

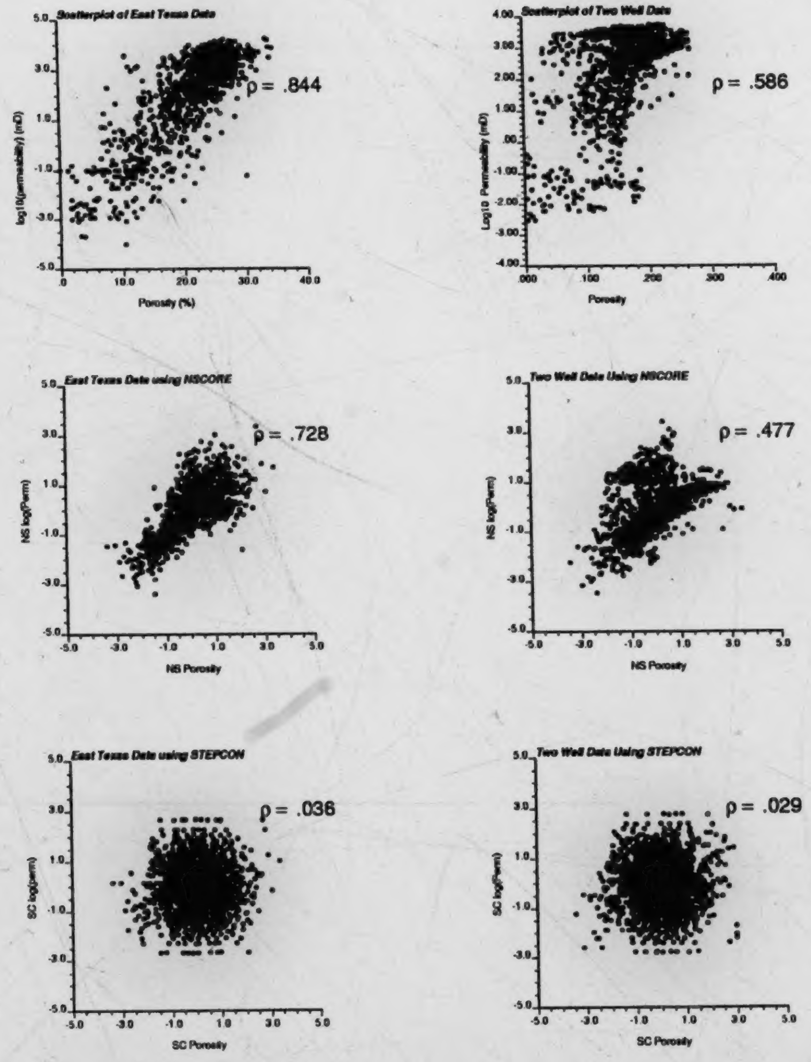


Figure 3.3: Illustration of cross plots of the original data (top), normally transformed data (middle) and the stepwise conditionally transformed data (bottom) for porosity and log(permeability) for East Texas data (left side) and the Two Well dataset (right side).

data, respectively). The transformation removes the correlation *between* classes, but does not account for the correlation within a class.

Once transformed, there are several approaches to multivariate geostatistical simulation of the stepwise conditional scores. As mentioned, the cross covariance at $\mathbf{h} = 0$ is zero by construction; however, the cross covariance at $\mathbf{h} > 0$ may not be zero. There are two possible options for fitting the cross covariance $C'_{ij}(\mathbf{h}), \mathbf{h} > 0$ at large scale:

1. Assume independence for all lag distances after confirmation by calculating an experimental cross variogram or cross covariance, that is

$$C(Y_i'(\mathbf{u}), Y_j'(\mathbf{u}+\mathbf{h})) = C'_{ij}(\mathbf{h}) = 0, \text{ for } i \neq j, \forall \mathbf{h}$$

2. Model $C'_{ij}(\mathbf{h})$ consistent with a valid linear model of coregionalization (LMC).

The first option is simplest. Calculation of $C'_{ij}(\mathbf{h})$, $i \neq j$ will identify if further modeling of the cross covariance is required due to significant departures from independence.

The significant advantage of this method is that complex multivariate distributions are transformed to the well behaved Gaussian distribution. For example, non linear, heteroscedastic and constraint features (see Figure 1.1) are automatically built into the transformation.

3.1 Model of Coregionalization

It is of theoretical interest to understand the model of coregionalization implicit to the stepwise conditional transformation, and the conditions under which the assumption that all cross covariances at all spatial scales are zero, $C'_{ij}(\mathbf{h}) = 0, i \neq j, \forall \mathbf{h}$ are appropriate.

Transformation by the stepwise conditional procedure leads to an implicit model of coregionalization. The model of coregionalization is embedded within the transformation and back transformation. The model of coregionalization can always be understood numerically via simulation and calculation. This may be important for complex situations; however, for certain simple cases, an analytical examination may provide insight to this implicit model.

The stepwise conditional transformation of the primary variable is identical to its normal score transform. As a result, the covariance structure of the primary variable is the same as the covariance calculated from the conventional normal scores, that is, $C'_{11}(\mathbf{h}) = C_{11}(\mathbf{h})$.

Conditional transformation of the secondary variable results in $Y_2' \neq Y_2$, hence the covariance structure of Y_2' is different from that of Y_2. To gain a better understanding of the differences between these two covariance structures, a small analytical exercise was carried out. Two points separated by a distance $\mathbf{h}$ were considered (see Figure 3.4). At each point, Y_1 and Y_2 data are available. Both the original variables, Y_1 and Y_2, are homoscedastic, multi-Gaussian variables with univariate

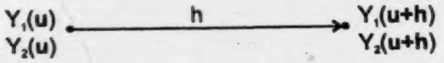


Figure 3.4: Schematic illustration of two points separated by a distance **h**.

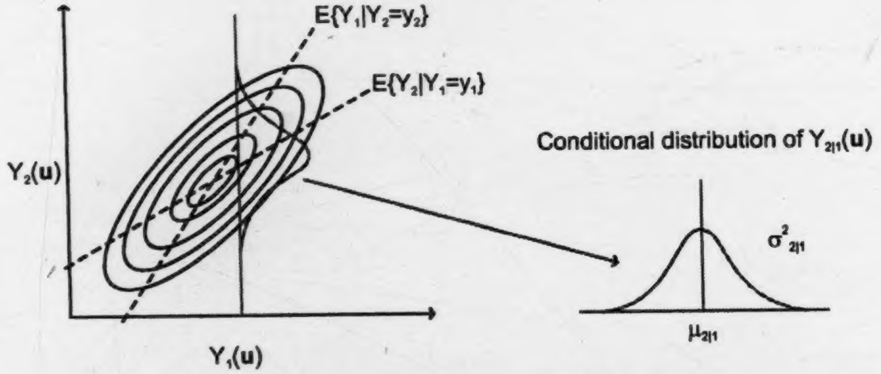


Figure 3.5: Conditional distribution of the secondary variable, $Y_2(\mathbf{u})$ with conditional mean, $\mu_{2|1}(\mathbf{u})$ and variance, $\sigma_{2|1}(\mathbf{u})$.

N(0,1) distribution. Further, the covariance structure between Y_1 and Y_2 is defined analytically by a valid linear model of coregionalization (LMC) (Section 2.1.4).

Transforming the secondary variable conditional to the primary variable results in the following transform expression,

$$Y_2'(\mathbf{u}) \quad = \quad \frac{Y_2(\mathbf{u}) - \mu_{2|1}(\mathbf{u})}{\sigma_{2|1}(\mathbf{u})} \tag{3.1}$$

where $\mu_{2|1}$ and $\sigma_{2|1}$ are the mean and standard deviation of the conditional distribution, respectively, of the conditionally transformed Y_2'. Figure 3.5 shows a schematic illustration of the conditional distribution of the secondary variable, $Y_2(\mathbf{u})$, given $Y_1(\mathbf{u})$. Note that the transform given in Equation 3.1 is simply a standardization of the ccdf of $Y_2(\mathbf{u})$ given $Y_1(\mathbf{u})$ to be standard normal (that is with zero mean and unit variance).

The covariance model of the transformed secondary variable, Y_2', is given:

$$\begin{aligned} C_{22}'(\mathbf{h}) &= E\{Y_2'(\mathbf{u}) \cdot Y_2'(\mathbf{u}+\mathbf{h})\} \\ &= E\left\{\left(\frac{Y_2(\mathbf{u}) - \mu_{2|1}(\mathbf{u})}{\sigma_{2|1}(\mathbf{u})}\right) \cdot \left(\frac{Y_2(\mathbf{u}+\mathbf{h}) - \mu_{2|1}(\mathbf{u}+\mathbf{h})}{\sigma_{2|1}(\mathbf{u}+\mathbf{h})}\right)\right\} \end{aligned} \tag{3.2}$$

To determine the covariance structure of the transformed secondary variable, C_{22}', the local conditional distributions must first be defined.

The mean and standard deviation of the conditional distribution, $\mu_{2|1}(\mathbf{u})$ and $\sigma_{2|1}(\mathbf{u})$, are known in the case of multiGaussian variables. In fact, all conditional distributions are Gaussian with known mean and standard deviation. The mean and variance are only needed to calculate the covariance structure and are given by solving the kriging system of equations; these are given by the kriged estimate and the error variance, respectively. Note that in this particular case, kriging is not being used in the conventional sense of estimation, but rather the system of equations yields the parameters of the ccdf *exactly* for the multiGaussian case [28, 40]. The simple kriging (SK) equations for this system are:

$$\mu_{2|1}(\mathbf{u}) = \sum_{\alpha=1}^{2} \lambda_\alpha Y_1(\mathbf{u}_\alpha) = \lambda Y_1(\mathbf{u}) + \lambda' Y_1(\mathbf{u}+\mathbf{h}) \quad (3.3)$$

$$\sigma^2_{2|1}(\mathbf{u}) = \sigma^2_E = C_{22}(\mathbf{0}) + \sum_{\alpha=1}^{2}\sum_{\beta=1}^{2} \lambda_\alpha \lambda_\beta C_{11}(\mathbf{u}_\alpha, \mathbf{u}_\beta) - 2\sum_{\alpha=1}^{2} \lambda_\alpha C_{12}(\mathbf{u}, \mathbf{u}_\alpha) \quad (3.4)$$

Minimization of the error variance (Equation 3.4) yields the normal equations:

$$\sum_{\alpha=1}^{2} \lambda_\alpha C_{11}(\mathbf{u}_\alpha, \mathbf{u}_\beta) = C_{12}(\mathbf{u}, \mathbf{u}_\alpha) \quad (3.5)$$

For the two points shown in Figure 3.4, the system of equations consists of

$$\lambda C_{11}(\mathbf{0}) + \lambda' C_{11}(\mathbf{h}) = C_{12}(\mathbf{0}) \quad (3.6)$$

$$\lambda C_{11}(\mathbf{h}) + \lambda' C_{11}(\mathbf{0}) = C_{12}(\mathbf{h}) \quad (3.7)$$

where λ is the weight given to $Y_1(\mathbf{u})$, λ' is the weight given to $Y_1(\mathbf{u}+\mathbf{h})$, $C_{11}(\mathbf{h})$ is the covariance between $Y_1(\mathbf{u})$ and $Y_1(\mathbf{u}+\mathbf{h})$, and $C_{12}(\mathbf{h})$ is the cross covariance between $Y_2(\mathbf{u})$ and $Y_1(\mathbf{u}+\mathbf{h})$. Note that for standard Gaussian variables, $C_{11}(\mathbf{0}) = 1$ and $C_{12}(\mathbf{0}) = \rho(\mathbf{0}) = \rho$. So Equations 3.6 and 3.7 becomes:

$$\lambda + \lambda' C_{11}(\mathbf{h}) = \rho \quad (3.8)$$

$$\lambda C_{11}(\mathbf{h}) + \lambda' = C_{12}(\mathbf{h}) \quad (3.9)$$

Solving equations 3.8 and 3.9 yields the following SK weights:

$$\lambda = \frac{\rho - C_{12}(\mathbf{h}) \cdot C_{11}(\mathbf{h})}{1 - C_{11}(\mathbf{h})^2} \quad (3.10)$$

$$\lambda' = \frac{C_{12}(\mathbf{h}) - \rho \cdot C_{11}(\mathbf{h})}{1 - C_{11}(\mathbf{h})^2} \quad (3.11)$$

These weights are then substituted into Equations 3.3 and 3.4 to obtain the mean and variance of the conditional distribution:

$$\mu_{2|1}(\mathbf{u}) = \lambda Y_1(\mathbf{u}) + \lambda' Y_1(\mathbf{u}+\mathbf{h})$$

$$\mu_{2|1}(\mathbf{u}) = \left(\frac{\rho - C_{12}(\mathbf{h}) \cdot C_{11}(\mathbf{h})}{1 - C^2_{11}(\mathbf{h})}\right) Y_1(\mathbf{u}) + \left(\frac{C_{12}(\mathbf{h}) - \rho \cdot C_{11}(\mathbf{h})}{1 - C^2_{11}(\mathbf{h})}\right) Y_1(\mathbf{u}+\mathbf{h}) \quad (3.12)$$

$$\begin{aligned}\sigma^2_{2|1}(\mathbf{u}) &= C_{22}(0) - \left\{\lambda C_{12}(0) + \lambda' C_{12}(\mathbf{h})\right\} \\ \sigma^2_{2|1}(\mathbf{u}) &= 1 - \left(\frac{\rho - C_{12}(\mathbf{h}) \cdot C_{11}(\mathbf{h})}{1 - C^2_{11}(\mathbf{h})}\right) C_{12}(0) - \left(\frac{C_{12}(\mathbf{h}) - \rho \cdot C_{11}(\mathbf{h})}{1 - C^2_{11}(\mathbf{h})}\right) C_{12}(\mathbf{h}) \quad (3.13)\end{aligned}$$

Similar expressions for the conditional mean, $\mu_{2|1}(\mathbf{u}+\mathbf{h})$, and standard deviation, $\sigma_{2|1}(\mathbf{u}+\mathbf{h})$, of $Y_2'(\mathbf{u}+\mathbf{h})$ can be determined using the two data for Y_1.

Substitution of $\mu_{2|1}$ and $\sigma_{2|1}$ into equation 3.2 shows that the covariance structure of the conditionally transformed variable, Y_2', implicitly incorporates the direct and cross-covariance structure of the original variables, Y_1 and Y_2. The new model of coregionalization implicitly invoked via the stepwise conditional transform is a function of the original variable covariance structures, that is,

$$\begin{aligned}C'_{11}(\mathbf{h}) &= C_{11}(\mathbf{h}) \\ C'_{12}(\mathbf{h}) &= g(C_{11}(\mathbf{h}), C_{12}(\mathbf{h}), C_{22}(\mathbf{h})) \\ C'_{22}(\mathbf{h}) &= f(C_{11}(\mathbf{h}), C_{12}(\mathbf{h}), C_{22}(\mathbf{h}))\end{aligned}$$

where f and g are different functions of the direct and cross covariance structure of the original variables. $C'_{12}(\mathbf{h})$ can be assumed to be zero, after numerical verification.

Intrinsic Coregionalization. For the special case of an intrinsic coregionalization, that is when $C_{22}(\mathbf{h}) = C_{11}(\mathbf{h})$ and $C_{12}(\mathbf{h}) = \rho_{12}(0) \cdot C_{11}(\mathbf{h}) = \rho \cdot C_{11}(\mathbf{h})$, $C'_{12}(\mathbf{h})$ is zero, the SK weights in Equations 3.8 and 3.9 become:

$$\begin{aligned}\lambda &= \rho \\ \lambda' &= 0\end{aligned}$$

The mean and variance of the conditional distribution reduce to

$$\begin{aligned}\mu_{2|1}(\mathbf{u}) &= \rho Y_1(\mathbf{u}) \\ \sigma^2_{2|1}(\mathbf{u}) &= 1 - \rho^2\end{aligned}$$

So the conditional distribution of Y_2' for the intrinsic case is $N(\rho Y_1(\mathbf{u}), 1 - \rho^2)$, which agrees with the conditional distribution obtained by applying Bayes law on the conditional expectation of two standard multiGaussian variables.

Using the mean and variance of the conditional distribution, the covariance model of the transformed variable can be determined:

$$\begin{aligned}
C'_{22}(\mathbf{h}) &= E\{Y'_2(\mathbf{u}) \cdot Y'_2(\mathbf{u}+\mathbf{h})\} \\
&= E\left\{\left(\frac{Y_2(\mathbf{u}) - \rho Y_1(\mathbf{u})}{\sqrt{1-\rho^2}}\right) \cdot \left(\frac{Y_2(\mathbf{u}+\mathbf{h}) - \rho Y_1(\mathbf{u}+\mathbf{h})}{\sqrt{1-\rho^2}}\right)\right\} \\
&= \frac{1}{1-\rho^2} E\left\{[Y_2(\mathbf{u}) - \rho Y_1(\mathbf{u})] \cdot [Y_2(\mathbf{u}+\mathbf{h}) - \rho Y_1(\mathbf{u}+\mathbf{h})]\right\} \\
&= \frac{1}{1-\rho^2}\left\{C_{22}(\mathbf{h}) - 2\rho C_{12}(\mathbf{h}) + \rho^2 C_{11}(\mathbf{h})\right\} \\
&= \frac{1}{1-\rho^2}\left\{C_{11}(\mathbf{h}) - 2\rho(\rho C_{11}(\mathbf{h})) + \rho^2 C_{11}(\mathbf{h})\right\} \\
&= \frac{1}{1-\rho^2}\left\{C_{11}(\mathbf{h}) - \rho^2 C_{11}(\mathbf{h})\right\} \\
&= \frac{C_{11}(\mathbf{h})(1-\rho^2)}{1-\rho^2} \\
&= C_{11}(\mathbf{h}) \qquad (3.14)
\end{aligned}$$

So the covariance structure for the transformed secondary variable, Y'_2 reduces to the covariance structure of the primary variable, Y_1. The cross covariance for the intrinsic coregionalization case, $C'_{12}(\mathbf{h})$, is zero for all distances. This theoretical result is validated in the numerical exercise below. Note that this is an "extreme" case of the model of coregionalization implicit to the stepwise conditional transformation.

Numerical Exercise. The result of applying this transformation is independence of the transformed variables at $\mathbf{h} = 0$, since each class of Y_2 data is independently transformed to a standard normal distribution. There is no guarantee of independence for distance lags greater than zero ($\mathbf{h} > 0$). The new model of coregionalization is complex. As shown above, the assumption of $C'ij(\mathbf{h}) = 0, \forall \mathbf{h}, i \neq j$ was only demonstrated for the case of an intrinsic coregionalization. To validate this theoretical result, a numerical exercise was performed involving two multi-Gaussian variables, Y_1 and Y_2, with the same direct isotropic variogram consisting of a combination of two spherical models with a range given by a:

$$\gamma(\mathbf{h}) = 0.5Sph_{a=3}(\mathbf{h}) + 0.5Sph_{a=15}(\mathbf{h})$$

The correlation between Y_1 and Y_2 was chosen to be 0.70. Three different cross variograms were considered: "short-range", "intrinsic", and "long-range". The "short-range" case gives maximum variance contribution to the short range structure; while the "long-range" case gives maximum variance contribution to the long range structure. Note that *maximum variance contribution* refers to the maximum contribution allowable under the LMC. The cross semivariogram models are given below and illustrated in Figure 3.6.

$$\begin{aligned}
&short-range: & \gamma(\mathbf{h}) &= 0.50Sph_{a=3}(\mathbf{h}) + 0.20Sph_{a=15}(\mathbf{h}) \\
&intrinsic: & \gamma(\mathbf{h}) &= 0.35Sph_{a=3}(\mathbf{h}) + 0.35Sph_{a=15}(\mathbf{h}) \\
&long-range: & \gamma(\mathbf{h}) &= 0.20Sph_{a=3}(\mathbf{h}) + 0.50Sph_{a=15}(\mathbf{h})
\end{aligned}$$

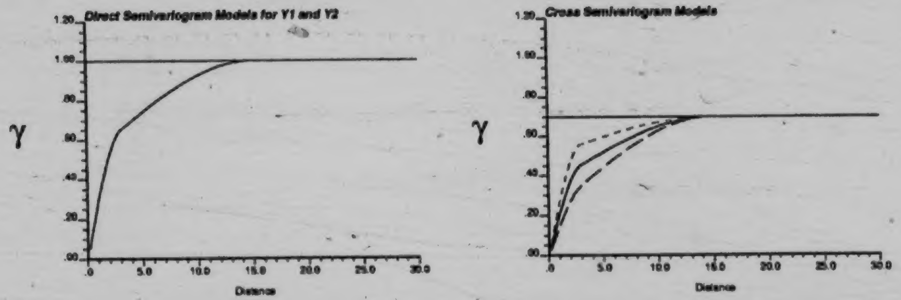


Figure 3.6: Direct semivariogram of Y_1 and Y_2 (left), and the three different cross semivariogram models (right) : short-range (top left, dash), intrinsic (middle, solid), and long-range (lower right, dash).

For each case, stepwise conditional transformation was applied, direct and cross variograms were calculated and modeled, sequential Gaussian simulation was performed, simulated values were back transformed, and the resulting simulated direct and cross variograms were examined. Figure 3.7 shows the direct variograms for Y_2' and the cross variogram of Y_1 and Y_2', following application of the stepwise transform.

In the short-range scenario, the cross variogram is slightly higher than zero over small lag distances and then returns to zero. Conversely, the long-range scenario shows that the cross variogram is negative over the short-range. For these two cases, fitting a cross semivariogram model to the structures is challenging; the thick solid line in each figure shows the dampened hole effect model that was fitted to the simulated results. Unlike the two extreme cases, the intrinsic case showed independence of the transformed pairs, with no deviation from zero over all lags. As predicted by theory, independence at $\mathbf{h} > 0$ is satisfied for the intrinsic case.

Comparison of the analytical result in Equation 3.14 with the numerical results shown in Figure 3.7 for the intrinsic case shows that the numerical result deviates only slightly from the analytical solution. This deviation can be attributed to ergodic fluctuations.

Following simulation, the values were back transformed and the cross variogram was checked for each scenario. Figure 3.8 shows the model cross variograms of the original variables and the average cross variogram obtained after simulation of the conditionally transformed variables. The range of correlation is approximately preserved, i.e., the short range model produces an average cross variogram with the shortest range of the three simulated scenarios. As well, the variogram structure of the extreme cases (short- and long-range cross variograms) are shifted towards the intrinsic model, since this is the model that has been implicitly assumed.

The direct variograms of the resulting secondary variable were also examined and showed a shift in the direct model towards the opposite extreme in order to produce an overall shift in the cross variogram towards the intrinsic case. Thus, in the short range cross variogram scenario, the direct variogram for the resulting simulated secondary variable showed that greater variance contribution was given to the long range structure. This yields an overall shift in the cross variogram to

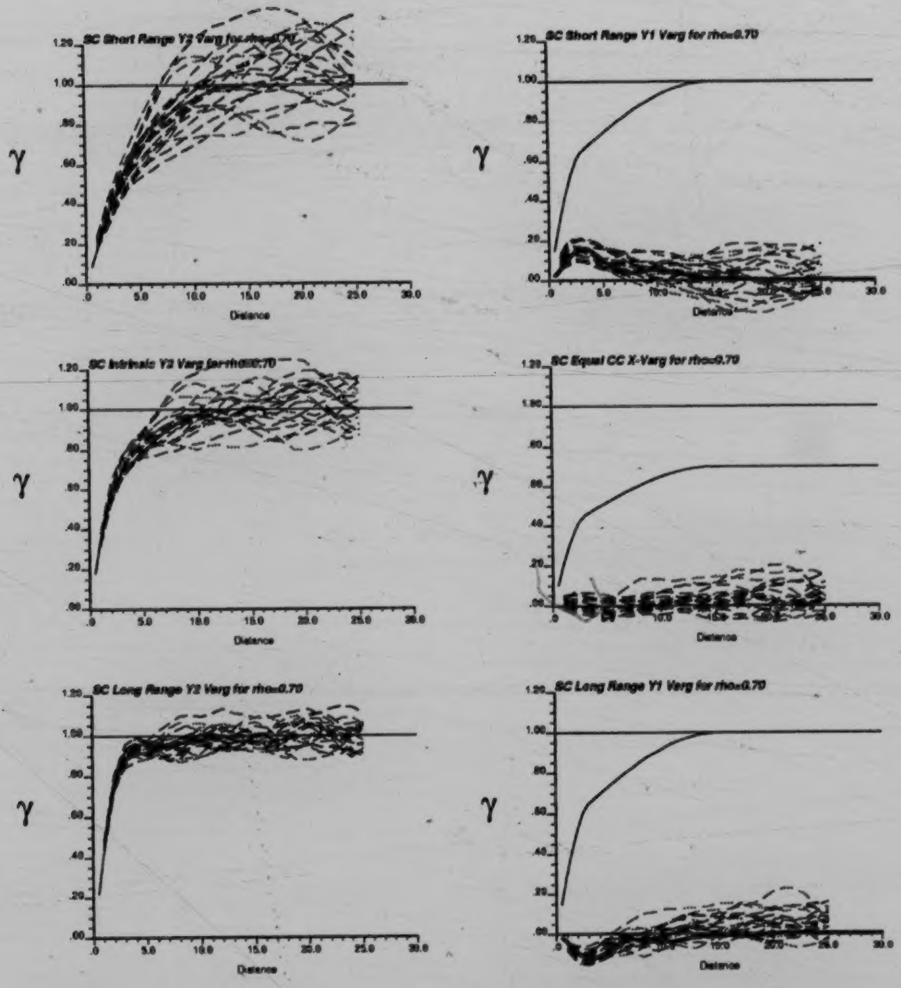


Figure 3.7: Direct semivariogram of Y_2' (left) and cross semivariogram of Y_1' and Y_2'(right), after stepwise conditional transformation for the short range (top), intrinsic (middle) and long range (bottom) scenarios. The (thin) solid black line on the cross semivariograms represent the cross semivariogram model used to create the unconditioned simulation prior to transformation; while the thick solid black line on the cross semivariograms show the dampened hole effect model that was fitted.

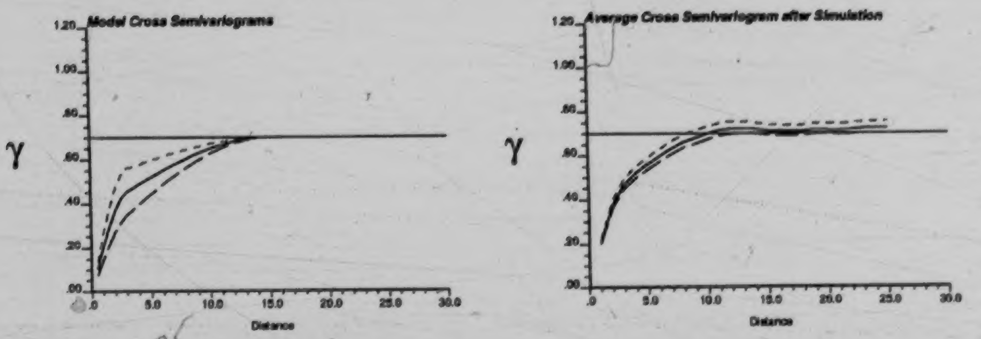


Figure 3.8: Input model of cross variogram of Y_1 and Y_2 (left), and the average cross variogram obtained after simulating with stepwise transformed variables Y_1' and Y_2' (right). In both cases, the variograms follow the same line code: short-range (top left, dash), intrinsic (middle, solid), and long-range (bottom right, dash).

be closer to the intrinsic case. Conversely, the long range cross variogram scenario resulted in a direct variogram that gave greater variance contribution to the short range structure. This results in a cross variogram that is shifted closer to the intrinsic model. Only the intrinsic case showed no shift in the direct variogram of the simulated secondary variable.

3.2 Links to "Cloud" Transform / P-field

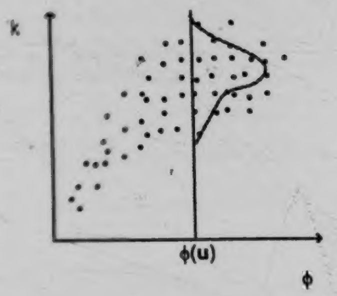
The stepwise conditional transformation bears some similarities to a "cloud" transform that is sometimes used in the modeling of petrophysical properties for reservoir characterization. Core porosity and permeability are usually available and typically used to establish the bivariate relationship between porosity and permeability. Log data are also typically available and are deemed to be more reliable for conditioning of 3-D models. Porosity is first simulated using the log data.

A 3-D model is also required for permeability. The cloud transform modeling of permeability proceeds in two steps: (1) generate a 3-D correlated random field of probabilities, and (2) draw a simulated value for permeability using its collocated porosity model value to determine the conditional distribution of permeability [5].

The first step in permeability modeling is essentially one part of the p-field simulation algorithm (Section 2.1.7). Recall that in p-field simulation, the parameters for the local conditional distributions are defined by kriging the 3-D grid using the available data. A probability field is generated for the same 3-D grid that has some spatial correlation. These probabilities are used to draw from the local conditional distributions that were previously obtained from kriging. In this way, the simulated values are spatially correlated and are also conditioned to the data. Similar to p-field simulation, a 3-D correlated random field of probabilities are required for the cloud transform:

- Determine the conditional distribution of permeability. This is based on the collocated porosity that is available from the initial step of modeling porosity,

1) Determine conditional distribution of permeability based on collocated porosity value.



2) Find the quantile of permeability associated to the probability from the p-field.

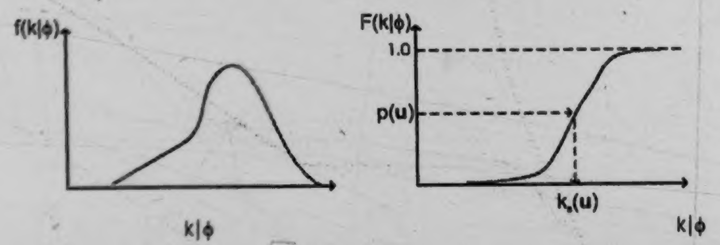


Figure 3.9: Schematic illustration of cloud transform for petrophysical modeling: (1) Based on porosity value at location **u**, determine the corresponding conditional distribution from the bivariate distribution of core porosity and permeability, (2) Draw a simulated value of permeability using the probability from the correlated random field of probabilities.

and the bivariate distribution of the core porosity and permeability (Step 1 in Figure 3.9).

- Draw a simulated value from the conditional distribution of permeability, using the probability at that location (from the p-field) (Step 2 in Figure 3.9).

The link between this "cloud" transform approach and the stepwise conditional transformation lies in the use of the global bivariate distribution to determine conditional distributions. In practice, the conditional distributions are constructed by binning the data. Where the cloud transform simply draws from the conditional distribution, the stepwise conditional transform approach does a normal score transformation of the same conditional distribution. Both methods will reproduce the bivariate distribution, but the stepwise transform approach has the potential for modeling permeability independent of porosity. This is contrasted with cloud transform of permeability where the p-field is independent of porosity, but conditioning this field to local data is not trivial (see Section 2.1.7).

3.3 Remarks

Conditional transformation of the data results in transformed secondary variables that are combinations of multiple "real" variables. Consequently, the associated covariance structure of the secondary transformed variables implicitly incorporates the direct and the cross covariance structure of the original variables. Thus, a new model of coregionalization is implicitly invoked in the covariance structure of all transformed secondary variables.

Although all discussions thus far have assumed that the multivariate data are continuous variables, this technique does not preclude application to categorical or discrete data. In fact, the stepwise transform can be performed on continuous, categorical, or a combination of both categorical and continuous data.

In practice, however, discrete data in the natural resources industry are geological codes. Similarly, well log samples may be coded with different categorical values to differentiate the lithofacies from which the samples were extracted. Data found within different rock types typically have different statistical, as well as structural, properties. Consequently, stationarity decisions are applicable only within rock types, and each rock type should be independently modeled to reflect its distribution and continuity structure. In this context, the stepwise conditional transform is not a favourable alternative to making appropriate stationarity decisions to subdivide the domain into more homogeneous subzones.

The next chapter will explore some of the practical details of implementing the technique. Specifically, the following issues will be addressed:

- Data related issues, including the effect of outliers and spikes, on the forward and back transformation.
- The requirement for sufficient data and the specification of number of classes are closely related issues.
- The proposed application of a smoothing algorithm to infer a reliable conditional distribution when there are insufficient data to proceed with the transform.
- The transformation order that should be adopted in order to preserve the spatial continuity of the original variables.

Chapter 4

Considerations for Real Data

There are a number of different considerations for the application of the stepwise conditional transform to real data. The following issues are important: (1) data-related issues, (2) number of data and classes, (3) inference of multivariate distributions in presence of sparse data, and (4) ordering of variables for transformation.

4.1 Data Related Issues

This section addresses potential errors in the data, the possibility of outliers, spikes, spatial clustering of the data and missing data.

Errors and Outliers. Sampling errors are common as a result of the data collection process and/or the recording of observed measurements. The result of such errors may lead to outliers such as anomalously high values in an otherwise low-valued distribution or vice versa.

A nickel laterite data will be used to study the effect of outliers on the simulation results of stepwise conditionally transformed variables [48]. Only two variables will be examined: Ni and Fe. This data was minimally cleaned by the removal of three outliers. In the back transformation stage of the workflow, simulated values are interpolated in the transformation table to return the simulated value to the units of the original data. Outliers extend the domain of this interpolation. As a result, the crossplots of back transformed values show a bivariate distribution that reproduces these outlier values with a number of interpolated values, which may be inappropriate. Figure 4.1 shows the comparison of the original data distribution (uncleaned and cleaned) and the resulting back transformed simulated values for Ni and Fe.

Note that in Figure 4.1, the simulated crossplot of the uncleaned data does not show *exact* reproduction of some outliers. In particular, the data pair with $Ni = 5.1$ and $Fe = 71.0$ is not reproduced. This is attributed to an implementation issue in sequential Gaussian simulation (SGS) in which the data are assigned to grid nodes to speed up the simulation. In these cases, if there is another data that is closer than the outlier to the centre of the same grid node, then the outlier data will not be assigned and hence will not be reproduced exactly.

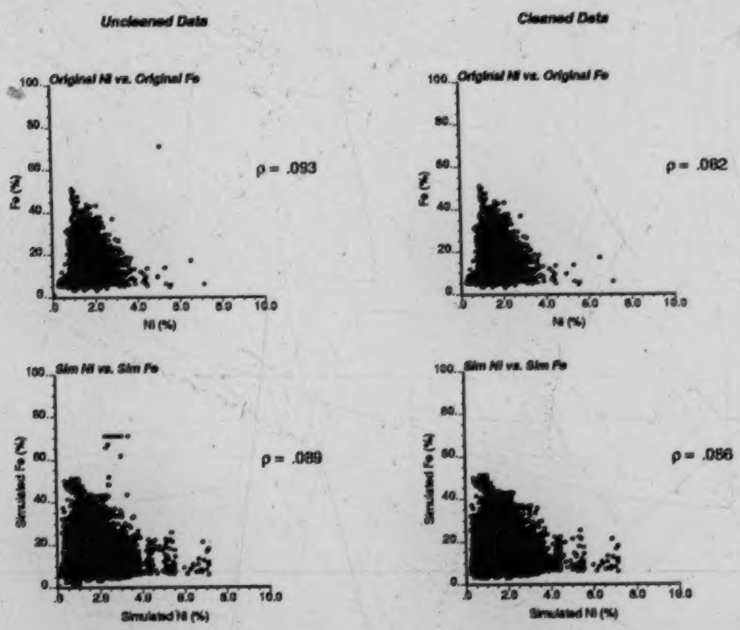


Figure 4.1: Effect of outliers on the reproduction of the bivariate distribution of a nickel laterite dataset. The crossplots of the uncleaned data (top left) and the cleaned data (top right) are compared with the crossplots of the corresponding simulated results in the bottom row.

Exploratory data analysis should reveal any outliers. In some cases outliers may be "legitimate" data, that is, these data may represent a true phenomena such as anomalously high gold grades in a vein type deposit. In other cases where outliers result from sampling errors, these problematic data should be removed prior to transformation to avoid simulating features that are not truly part of the mineralization.

Spike of Constant Values. It is common to encounter some proportion of values that are constant. In mining, this constant value may be attributed to low values below some detection limit. Most literature refers to this effect as either a spike [20, 73, 75], the zero effect [13, 40], or an atom at the origin of a histogram [13].

The presence of a spike results in a quantile transform that is not unique [20]; despiking becomes an important issue. This issue is important for any quantile transform [13]. Figure 4.2 shows a schematic illustration of the basic idea behind "despiking" or "breaking the ties".

Random despiking is the simplest solution [20], which simply involves adding a small random component to the constant values and sorting them accordingly. Verly proposed a despiking approach by taking local averages around the constant values, and ranking them based on these averages [75, 73].

In this aspect, the main concern for the stepwise conditional transform is the

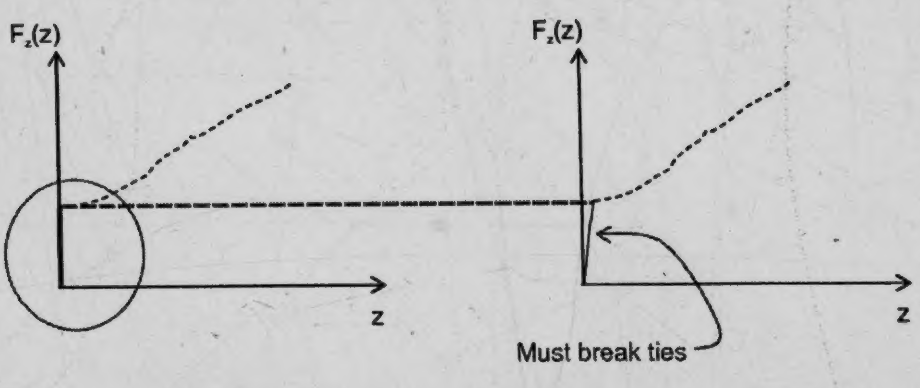


Figure 4.2: Schematic illustration of tie breaking to despike a distribution for quantile transformation.

presence of a large proportion of spikes at the border of two classes. For example, if the proportion of zeroes is 27 % and the probability threshold between the first and second class is 0.20, then the issue arises as to which 7% of the 27% constant values should be categorized to the second probability class (see Figure 4.3).

Back transformation is sensitive to despiking. In order to reproduce the conditional distributions, the constant values must be back transformed to the same class as originally assigned in the forward transformation. This may occur if the transformation and back transformation are performed outside of the simulation program. Experience has shown that in these instances, numerical precision in the transformation table may damage reproduction. Ideally, the transform should be integrated into the simulation algorithm and performed internally so as to maintain consistency in precision and hence class assignment.

Clustering and Non-Representative Data. It is common in natural resources exploration to take more samples in interesting or important areas. Depending on the attribute of interest, these areas may correspond to high or low valued regions. The equal weighted statistics of these clustered samples are not representative of the population to be modeled.

Declustering tools, such as polygonal and cell declustering, are applied to mitigate the effect of preferential sampling on the histogram and summary statistics [20]. The resulting histogram is simply the original data with weights that differ from the original equally weighted distribution. Note that declustering does not remove outliers from the dataset, the sample may be given lower weights, but the outlier is still present and will have an effect on the transformation. This declustered distribution often becomes the reference or target distribution to be reproduced by simulation.

Since a quantile transform can be applied on any univariate distribution, the transformation itself cannot correct for a non-representative distribution as a result of clustering. Consequently, declustering should be a consideration prior to appli-

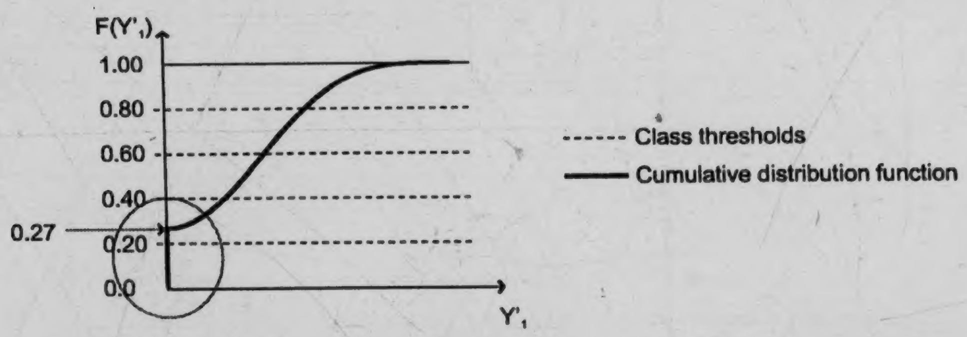


Figure 4.3: Despiking issue associated to class thresholds for stepwise transformation. This occurs when the proportion of spikes or constant values (e.g. 27% or 0.27) overlaps a class threshold (e.g. 20%) that differentiates between conditional distributions.

cation of this and any other quantile-based transformation methods. The stepwise conditional transformation should be applied on the distributions that are deemed to be representative of the domain to be modeled.

4.2 Number of Data and Class Specification

There must be sufficient data to reliably identify each conditional distribution used for transformation. Otherwise, the transformation may be unreliable and artifacts could be introduced in the back transformation.

The reliability of a distribution depends on the available number of samples, that is, a distribution defined by a large number of data is considered to be more reliable than a distribution that is based on few samples. For stepwise transformation, the number of data required increases as a power of the number of variables to transform. There is no general rule, however, 10^N to 20^N data, where N is the number of variables, would permit each distribution to be discretized into 10-20 classes with a minimum of 10-20 data [44].

Number of Classes. As the number of classes increase, the number of data per class will decrease. To illustrate the sensitivity of the transformation to the number of classes, consider porosity and permeability taken from the "Two Well" dataset from Chapter 1. There are approximately 1800 well log samples available. For Gaussian variables, zero correlation is a sufficient condition for independence. Depending on the available data and the number of classes specified, the tendency of the correlation coefficient towards zero will be examined. Correlation coefficients for number of classes from 1 to 10 and 15 to 50 in increments of 5 were calculated. Figure 4.4 shows the relation between the number of classes and the correlation coefficient. We can see that the correlation tends to zero as the number of class increases, flattening off at around 10 classes.

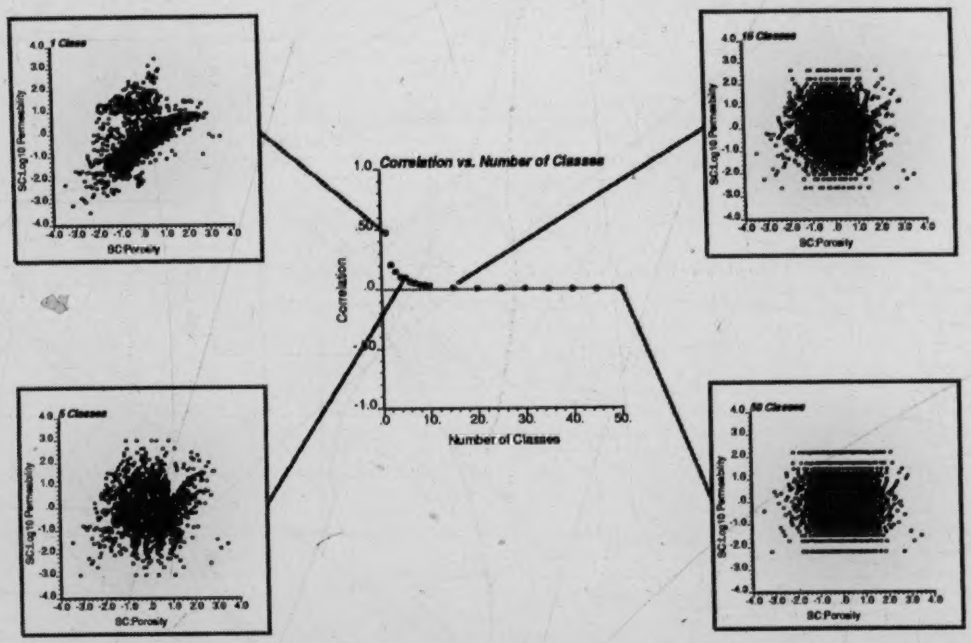


Figure 4.4: Change in correlation between the stepwise conditionally transformed variables as the number of probability classes changes for the Two Well data. Cross plots for the SC variables when the number of class equals 1, 5, 15 and 50 are shown. There are approximately 1800 original paired data.

Note that specifying one class for transformation is identical to performing independent normal score transformation for each variable. Thus the correlation for one class is the same as the correlation between the normal scores of the two variables.

As well, some non-linear features remain evident in some of the crossplots after transformation. In particular, the crossplots corresponding to 5 and 15 classes show slight departures from the independent bivariate Gaussian distribution at high and low values. Although the stepwise conditional transformation removes correlation between the transformed variables, it does so by removing correlation *between* the classes. There is no control on the correlations *within* the classes, and so these departures may result.

Furthermore, it is evident that a banding effect becomes more apparent as the number of classes increase. This results from the number of data found within a conditional distribution, that is, as more classes are specified, fewer data can be found within each class and approximately the same number of data are in each class. Recall from the normal score transform (Section 2.1), that if the number of data within each class is the same, then the normal scores of one class to the next are exactly the same (see Equation 2.5) [48].

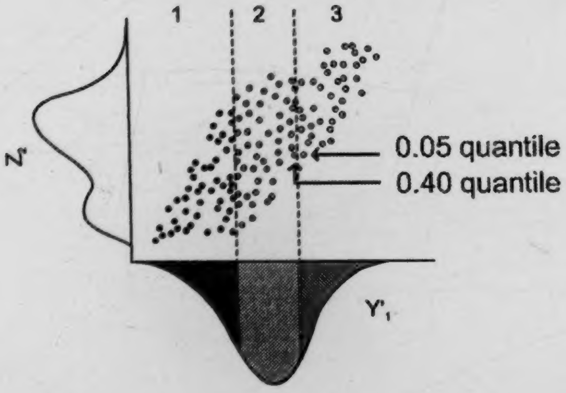


Figure 4.5: Differences in transformed secondary data due to class partition: the same original value for Z_2 may have different transformed values depending on the partitioning of the primary variable, Y_1' to obtain the conditional distributions to be transformed.

Determination of Class Partitions. The classification of data is based on partitioning the standard normal distribution. The thresholds are chosen to correspond to equal probability intervals. Depending on the primary data, the conditional transformation of secondary variables may result in the presence of artifacts in the transformed distribution. For example, two identical secondary data values should have the same transformed values (Y_2'). However, if the corresponding primary data belong to different probability intervals, then transformation may produce significant differences in the secondary variable Y_2' (Figure 4.5).

There is no requirement for the classes to be established based on equal probability intervals; they can also be calculated based on equal data value intervals or even user-specified intervals. For these two cases, the practitioner may encounter some artefacts as a result of insufficient data to define a specific conditional class, especially in the case of skewed distributions.

Dynamic Class Expansion. To mitigate the effect of a poorly defined conditional distribution due to few data, an option is to allow dynamic class expansion. This expansion occurs when a minimum number of data are not found within the class thresholds.

Figure 4.6 shows dynamic class expansion as it applies to the transformation of two and three variables; the method works for any number of variables. The current implementation expands the class by half an interval on either side of the two threshold values for each variable, effectively increasing the class size to twice the size of one regular class. This process occurs iteratively until the minimum number of data are found. In practice, the class expansion is typically only applied on the third or fourth variable, depending on data availability. Reducing the expansion to a smaller fraction of the class size would avoid large overlaps between classes; this is straightforward to implement and costs little in terms of computational effort.

Note that the purpose of class expansion is to identify a more reliable conditional distribution for the class. The only data that are transformed using this reference distribution are the actual data found within the initial class thresholds. The next section discusses an alternative to dynamic class expansion which is to smooth the multivariate distribution to allow transformation with sparse data.

4.3 Transformation in Presence of Sparse Data

There must be sufficient data to identify all conditional distributions in the stepwise transformation. Sparse data leads to erratic and nonrepresentative conditional distributions. Sparse data could be supplemented by a smoothing algorithm to "fill-in" gaps in the raw-data multivariate distribution [45]. This is an alternative to the class expansion approach presented in Section 4.2.

Smoothing using kernel densities is robust and flexible [63]:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{4.1}$$

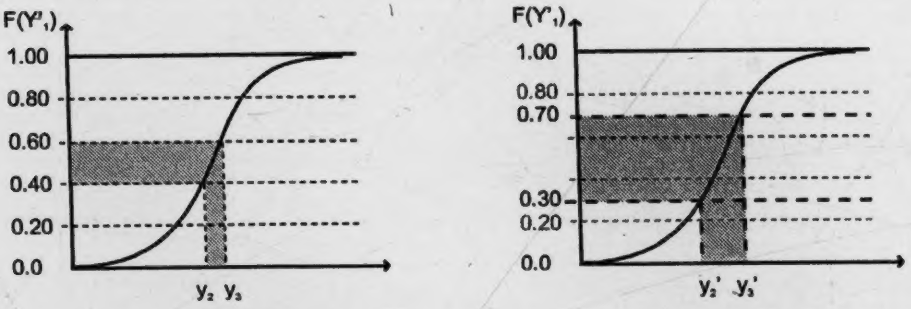
where n is the number of data, h is the bin width obtained by partitioning the range of the data (that is, between the minimum and maximum observed values) [34], $K(\cdot)$ is a kernel function associated to some specified density function. Since we are primarily concerned with discretizing the bivariate distribution, the kernel density is chosen to be a non-standard bivariate Gaussian density distribution with specified correlation:

$$f_{xy} = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \cdot e^{\frac{-1}{2(1-\rho^2)}\cdot\left(\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2\rho(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2}\right)}$$

where m_x and m_y are assigned the paired data values, σ_x^2 and σ_y^2 are user-specified variances associated to the two variables, and ρ is the correlation coefficient. Note that the above density function is the bivariate representation of Equation 2.40.

The general approach is to discretize the 2D space of a bivariate distribution into a grid to be populated with frequencies. A bivariate density distribution centered about each data pair is generated. The result is a "cloud" of values centered about the data . The size of this "cloud" is based on the variance specified by the user; for practical purposes, the variance can be determined empirically (typically 0.05 to 0.15). Further, the correlation coefficient of the kernel densities is typically set to the global correlation of the normal scores transform of the variables (since the kernels are chosen to be multivariate Gaussian). Figure 4.7 shows the effect of these two parameters on the bivariate kernel that is generated at a data point. The calculated frequencies are then averaged to obtain density estimates at each location within the grid. Discretizing the bivariate distribution for the stepwise conditional transformation will then be accomplished using this smoothed bivariate distribution. The basic steps in smoothing using a kernel estimator with a user specified correlation coefficient, ρ, and variance for each variable, σ_x and σ_y, are as follows:

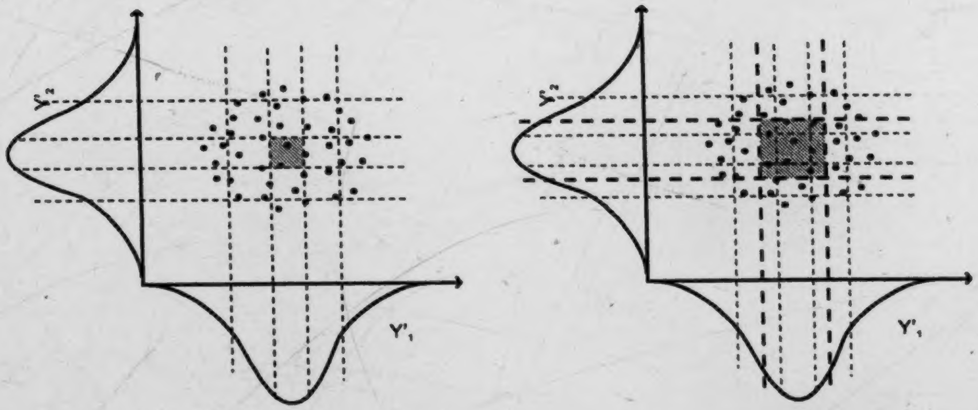
a) For two variables, expand classes of primary variable.



Original class thresholds

Expanded class thresholds

b) For three variables, expand classes of both primary and secondary variable.



Original class thresholds

Expanded class thresholds

Figure 4.6: Schematic Illustration of dynamic class expansion for equal probability partitioning to obtain minimum number of data: (a) bivariate case only requires expansion of probability class of the primary variable, (b) trivariate case, transformation of third variable requires class expansion of both the primary and secondary variable. In both cases, the size of the class is shaded in gray (original class definition (left) and expanded class (right)).

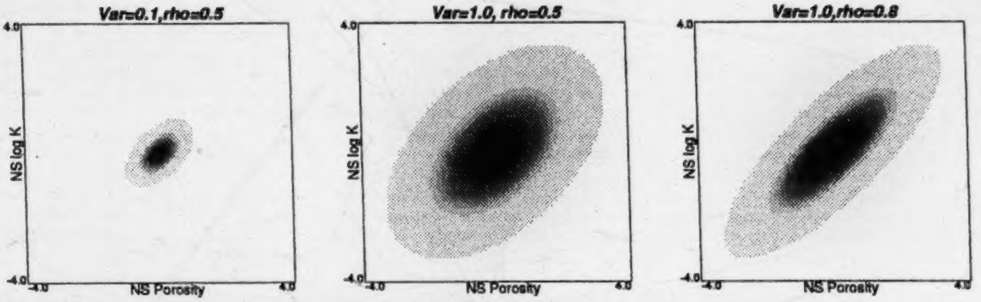


Figure 4.7: Effect of correlation and variance on bivariate kernel generated centered about a data pair. The first two plots (left and middle) show the effect of increasing variance, while the last two plots (middle and right) show the effect of changing the correlation coefficient. Note that a larger variance generates a larger "cloud" and allows for greater smoothing, while a larger correlation makes this "cloud" more linear.

1. Using the scatterplot limits for both variables, discretize the scatterplot to create a regular grid of X and Y values.
2. Go to each data pair:
   - Set $m_x = x$ and $m_y = y$.
   - Visit each node in the new scatterplot grid and calculate the bivariate frequency using the non-standard Gaussian density function.
3. Average all the calculated frequencies at each node.

The data should first be transformed into normal scores. Using the normal score values of the multivariate data, we smooth the bivariate distribution of the normal scores, then perform the stepwise conditional transformation on the original data and the smoothed distribution. Independent simulation of the model variables can now proceed in Gaussian space. Back transformation of the simulated values is implemented by calling on the univariate and the multivariate transformation tables.

This methodology was applied to a small petroleum related dataset consisting of only 27 data pairs of porosity and log(permeability). The correlation of the normal scores porosity and normal scores log(permeability) is 0.857, so this was the correlation specified for the kernels. The variance of the kernels was selected as 0.05 for both porosity and log(permeability), $\sigma_\phi^2 = \sigma_{logK}^2 = 0.05$. Figure 4.8 shows several comparative cross plots. The two cross plots of the stepwise conditionally transformed variables resulting from (1) only the data, and (2) the smoothed distribution have similar correlation magnitudes, but with opposite signs. Simulation and back transformation of the transformed variables according to the smoothed distribution shows good reproduction of the bivariate distribution. The banding effect that is visible in this crossplot is a consequence of back transforming values within

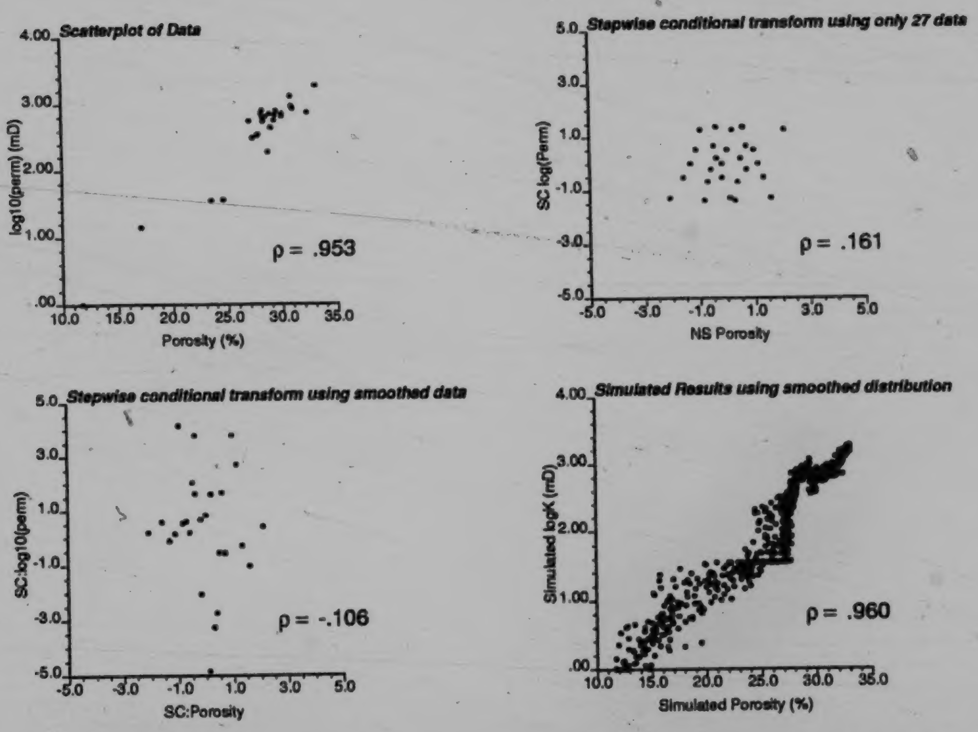


Figure 4.8: Small petroleum dataset consisting of only 27 samples. Cross plot of the original data (top left), cross plot of the stepwise conditionally transformed data using only the original 27 data values (top right), cross plot of stepwise conditionally transformed data using the smoothed distribution (bottom left), and a cross plot of the simulated values after back transformation (bottom right).

a sparsely defined class - smoothing in this instance does not fully compensate for defining a class with only two data points. The choice of a larger kernel would be required with the inevitable tradeoff of more smoothing (see Figure 4.7) [46].

The challenge of sparse data is not a limitation of the stepwise conditional transformation; all multivariate techniques require data. The limitation of working with isotopic sampling (Section 4.4), however, could preclude use of this transformation procedure.

4.4 Effect of Ordering

Consider two variables Z_1 and Z_2. Two possible scenarios exist for transformation: (1) choose Z_1 as primary variable and normal score transform to get Y_1, and then transform Z_2 to get $Y_{2|1}$; or (2) choose Z_2 as the primary variable to get Y_2, and then Z_1 is transformed to produce $Y_{1|2}$. Note the slight change in notation here from that in Chapter 3, the purpose is to differentiate between the primary variable and the conditionally transformed secondary variable. In this instance, the prime superscript ($'$) is missing from the Y variables; however, the conditioning is denoted by the subscripts and all Y variables in this Section refer to stepwise conditionally transformed variables.

In case (1) above, the simulation results for Y_1 would be identical to those obtained by conventional normal scores transformation of Z_1, and the same can be said for Y_2 in the second scenario. Simulation of the secondary variables, either $Y_{2|1}$ or $Y_{1|2}$, does not produce the same results as conventional simulation. The variogram of the secondary variable is a combination of the spatial structure of both original variables and the cross correlation of the two. In Section 3.1, it was shown that the model of coregionalization resulting from the stepwise transformation is complex to define analytically. Consequently, a numerical exercise was carried out to examine the difference in the variogram structure as a result of transformation ordering.

Unfortunately, this comparison is complicated by the fact that the variogram of the stepwise transformed secondary variable is not directly comparable to the conventional normal scores of the same variable. Figure 4.9 shows the following methodology undertaken for each sequence so that the variogram models that were compared for each variable were both the conventional normal scores variogram. For both ordering sequences, the variogram was calculated for both the stepwise transformed primary and secondary variable. Sequential Gaussian simulation was independently performed and back transformation returned the simulated values to the original units. Then, the simulated results were normal score transformed, and the corresponding variograms were determined. A comparison of the normal scores variograms for the same variable, when it is taken as (1) the primary variable and (2) the secondary variable, will show the effect of ordering.

This methodology was applied to the "Two Well" and "East Texas" data. Porosity and permeability were the two variables of interest. The first transformation order takes porosity as the primary variable, and the second takes permeability as the primary variable.

Figure 4.10 shows the comparison of the variograms for both ordering sequences of the Two Well dataset. The variograms for porosity show that when porosity was chosen as the primary variable, the post-simulation variograms closely follow the input normal scores variogram - as they should. Conversely, the variograms corresponding to the scenario in which porosity was the secondary variable shows greater variability and a shorter range. Differences in the permeability variograms as a result of transformation ordering sequence are not so obvious; however, the secondary variograms for permeability have longer range.

Figure 4.11 shows the comparison of the variograms for the East Texas data. Similar to the previous example, each scenario of ordering clearly shows departure of the secondary variable variograms from the direct variograms using the traditional normal scores. Unlike the Two Well example, the permeability variograms differ considerably after stepwise transformation. Further investigation showed that the stepwise transformation produced a secondary variable with higher nugget effect and longer range of correlation.

Recall that in Section 3.1, shifts in both the direct and cross variograms of the simulated transformed variables were examined in a numerical exercise. This exercise showed that cross variogram models that deviated from the intrinsic coregionalization model would shift the cross variogram model to be closer to the intrinsic

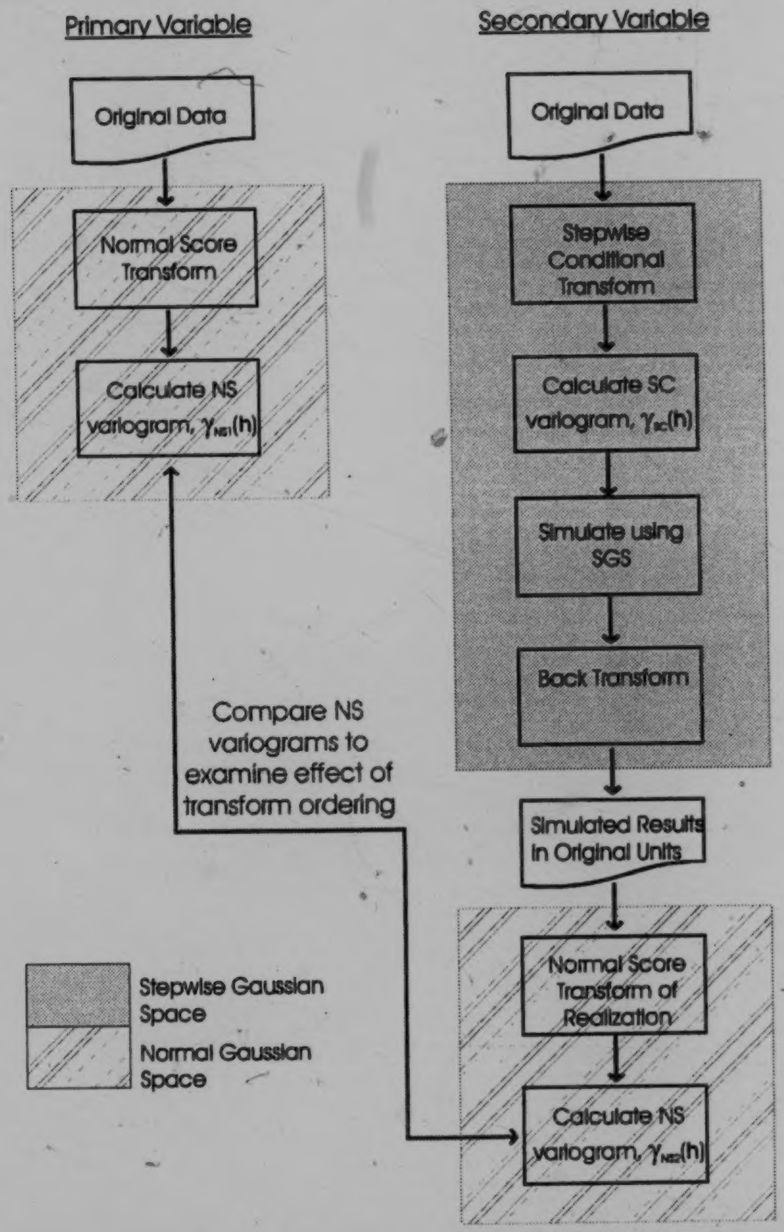

Primary Variable
Secondary Variable
Original Data
Original Data
Normal Score Transform
Stepwise Conditional Transform
Calculate NS variogram, γNS1(h)
Calculate SC variogram, γSC(h)
Simulate using SGS
Back Transform
Compare NS variograms to examine effect of transform ordering
Simulated Results in Original Units
Stepwise Gaussian Space
Normal Gaussian Space
Normal Score Transform of Realization
Calculate NS variogram, γNS2(h)

Figure 4.9: Methodology to examine the effect of transformation ordering on spatial structures for one variable. The flow chart shows the methodology to examine the spatial structure of one variable when it was taken as the primary variable (left), and when it was taken as the secondary variable (right), which was transformed conditional to a primary variable. The normal scores variograms, $\gamma_{NS1}(\mathbf{h})$ and $\gamma_{NS2}(\mathbf{h})$, were comparable since both were in the conventional Gaussian space. Note that although the stepwise and normal space are both Gaussian spaces, they reflect different transformations (legend, bottom left) .

model. This consequently results in deviations of the secondary transformed variable in order to give a cross variogram structure that is closer to the intrinsic case. Following these observations, the practitioner should first assess whether the structures of the direct variograms for the variables are similar (in type of structure, range and variance contributions). If the direct variograms are similar, then he should determine which direct variogram structure yields a closer fit to the cross-variogram structure when it is scaled by the cross correlation coefficient. The variable that gives the best fit should be chosen as the primary variable. This is consistent with the example using the Two Well dataset.

In the case where the direct variograms for the original variables are dissimilar, as in the East Texas data, the most continuous variable should be chosen as the primary variable. Continuity of a variable can be assessed by comparing the nugget effect, structure type and range of the variogram. Numerical examples show that variogram mismatch was minimized by this choice. The variograms for the second (and higher) variable often have a higher nugget effect than its corresponding normal scores variogram.

Nonisotopic Sampling. All discussions thus far have implicitly assumed that all data variables are available at all data locations; a situation called *isotopic* sampling. Unfortunately, in practice this assumption may be unrealistic; nonisotopic samples are common in resource valuation. In mining, nonisotopic data may arise as a result of two different sampling campaigns (one using diamond drill holes and another using reverse circulation drill holes), blast hole samples, and geophysical or seismic data [77]. In conventional geostatistics, this situation prevents inference of the cross covariance, in particular, the correlation at $\mathbf{h} = 0$ cannot be determined directly [77]. Essentially, this situation amounts to inaccessibility of the multivariate distribution at $\mathbf{h} = 0$.

In turn, this presents a serious limitation of the stepwise transformation. There are two particular issues associated to the presence of nonisotopic samples. Firstly, if the multivariate distribution is completely inaccessible (that is, there are no collocated data pairs), then the transform cannot be applied since there is no multivariate distribution to partition for transformation. Secondly, consider the situation where the multivariate distribution is only partially informed (that is, there are some isotopic samples), applying a multivariate transform to this data amounts to transforming a distribution that may be non-representative.

Notwithstanding the issue of non-representativity, the latter scenario presents a couple of possibilities for stepwise transformation. Consider locations where there are data for Z_j and no data for Z_i where $n_i < n_j$ and n is the number of data. The transformation of variable Z_j depends on prior transformation of Z_i; therefore, at locations where there are no Z_i, those Z_j data cannot be transformed. A straightforward solution is to choose the more densely sampled variable as the primary variable; however, the practitioner may decide that the more sparsely sampled variable is more important and so preservation of its spatial correlation is paramount. Alternatively, the chosen or preferred variable can be transformed and simulated at all locations. Then, the simulated primary variable can be used for later variables. Of course, there is no unique transformed value for secondary data at locations of

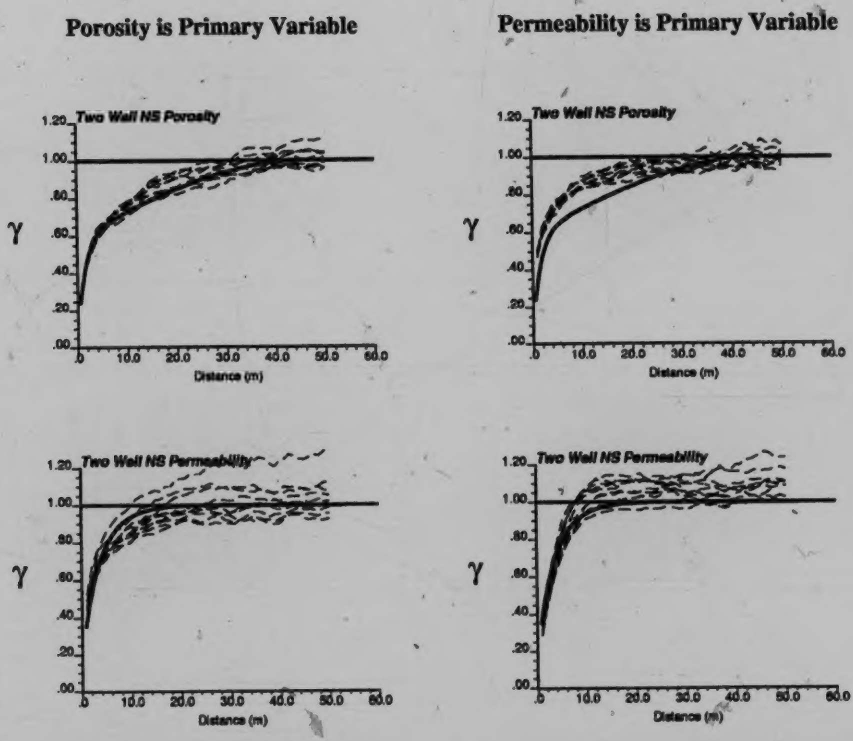


Figure 4.10: Effect of ordering using Two Well Data: normal scores variogram using simulated data for porosity (top) and permeability (bottom). In first scenario, porosity is taken as primary variable (left), and in the second scenario, permeability is chosen as the primary variable (right). In all cases, the thick solid line is the normal scores variogram model, the dashed lines correspond to the variogram of the simulated variable. Porosity is more continuous than permeability, and the greatest mismatch occurs when porosity is taken as the secondary variable.

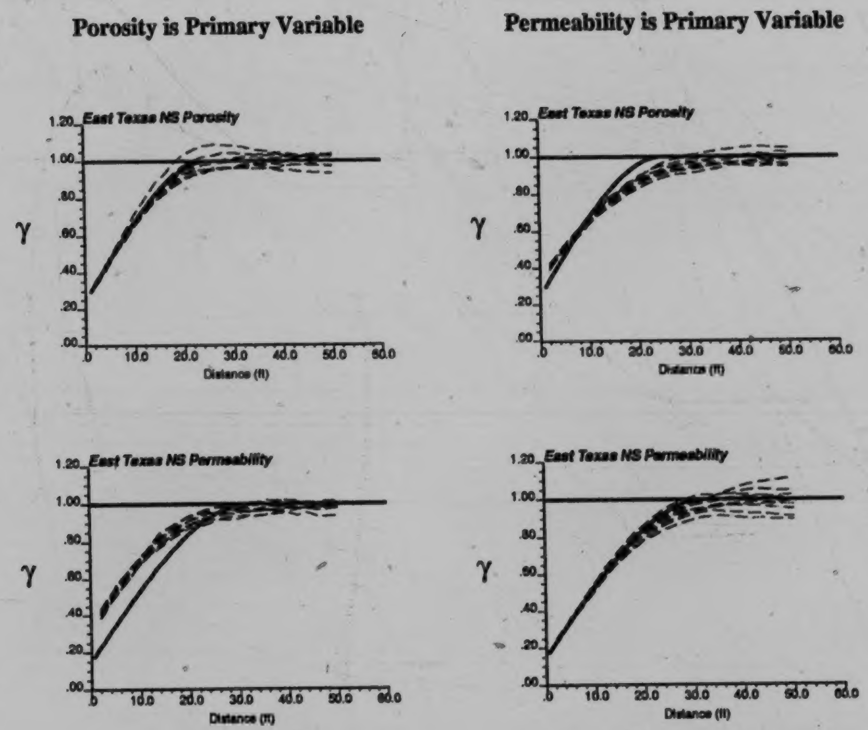


Figure 4.11: Effect of ordering using East Texas data: normal scores semivariogram using simulated data for porosity (top) and permeability (bottom). In first scenario, porosity is taken as primary variable (left), and in the second scenario, permeability is chosen as the primary variable (right). In all cases, the thick solid line is the normal scores semivariogram model, the dashed lines correspond to the variogram of the simulated variable. Permeability is more continuous than porosity, and the most significant mismatch in the semivariogram models occur when it is taken as the secondary variable.

nonisotopic sampling. This makes data analysis and inference of the variogram of secondary data difficult.

Application of the transform is not a problem in the presence of any number of exhaustive secondary information. Although this situation is technically the case of nonisotopic samples, it is distinct in that the discussion above mainly referred to the absence of collocated *hard* data. Exhaustively available secondary data are considered *soft* data. In these instances, the exhaustive secondary data should be chosen as the primary variable, and the hard data should be transformed conditional to the collocated secondary data.

4.5 Remarks

The stepwise conditional tranformation removes the correlation between the variables producing independent model variables at $\mathbf{h} = 0$. Cosimulation can proceed in one of two ways: (1) assume that $C'ij(\mathbf{h}) \simeq 0, i \neq j$ for $\mathbf{h} > 0$, or (2) model the multiple variograms consistent with LMC. The former case simplifies the cumbersome cosimulation process to independent simulation of the transformed variables; however, if in the latter case where $C_{ij}(\mathbf{h}) \neq 0, \mathbf{h} > 0$, then full cokriging is suggested. Note that modeling with the LMC may be challenging given that the correlation at $\mathbf{h} = 0$ is zero (or the sill is zero).

The correlation between the variables is injected during back transformation. This is a big advantage of transforming multiple variables in a stepwise conditional fashion. Any adverse effects of simulating non-multivariate Gaussian variables are mitigated by ensuring that the multivariate distributions of the transformed variables are truly Gaussian.

The covariance structure of the conditionally transformed secondary variables is a function of the direct and cross covariance model between the original variables. The effect of transformation ordering is observable in the departure of the semivariogram of the transformed variable from the original variable. This departure can be minimized by firstly determining if the direct variograms are similar. If so, then choose the variable that, when scaled by the correlation coefficient, provides the best fit to the cross variogram structure; otherwise choose the most continuous variable as the primary variable for stepwise transformation. In the presence of sparse data, dynamic class expansion can be allowed or a smoothing algorithm can be applied to model the conditional distributions based on the available data. The main limitation of the technique is the absence of any isotopic samples, which would preclude the use of this transformation altogether.

Application of the stepwise conditional transform results in a simplified work flow for simulation of multiple variables:

1. Perform stepwise conditional transformation for multiple variables.
2. Calculate and model variograms for each of the transformed variables, and verify that the cross covariance between the variables is close to zero for all lag distances, i.e. $C'ij(\mathbf{h}) \simeq 0, i \neq j$ for $\mathbf{h} > 0$.
3. Simulate each variable independently via Gaussian simulation.

4. Back transform in a stepwise conditional manner to obtain simulated values in original units.
5. Check model results: univariate and multivariate distributions, and variograms.

Figure 4.12 shows the results of following the above work flow to the Two Well data. Porosity and permeability are transformed in a stepwise conditional manner. The transformed variables are modeled and simulated independently. The two crossplots in the middle show the bivariate relationship before and after back transformation. The crossplot on the left is consistent with the independent simulation approach, while the crossplot on the right shows the reproduction of the bivariate relationship after back transformation.

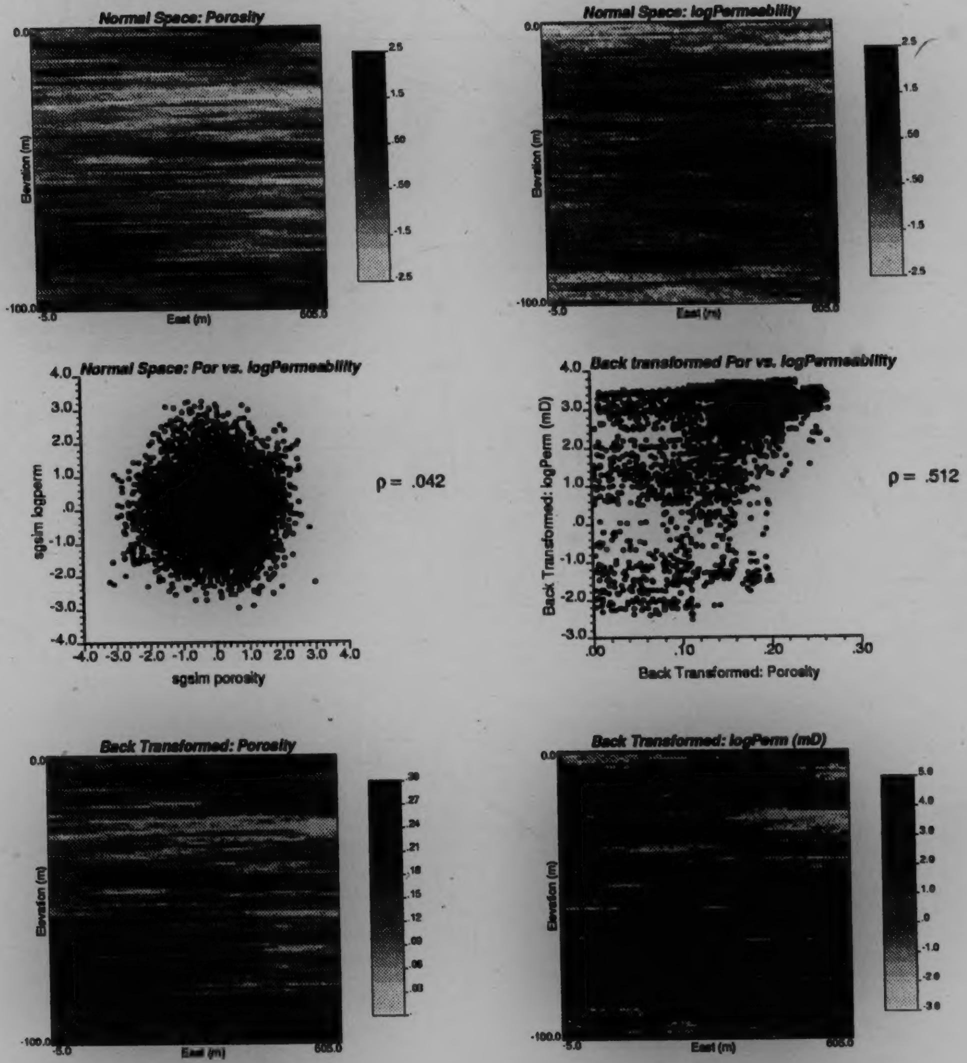


Figure 4.12: Example of independent simulation and back transformation of porosity and log permeability for Two Well Data: A cross section of one realization for porosity (left) and log permeability (right) in normal space (top row), crossplot of simulated values in normal space (middle left) and in original space after back transformation (middle right), and corresponding cross section of simulated porosity and log permeability in original space (bottom row). Cross plot reproduction can be compared to top right crossplot in Figure 3.3.

Chapter 5

Case Study: Multivariate Simulation of Red Dog Mine, Alaska, USA

The Red Dog mine is the world's largest Zn producer located 90 miles north of Kotzebue, Alaska, USA, owned and operated by Teck Cominco Limited. The deposit consists of sulphide ore zones in sedimentary exhalative (sedex) deposits, and is characterized by the presence of multiple metals and multiple ore types. The mine assays for as many as ten variables; the four primary ones being Zn, Pb, Fe and Ba.

A key issue is the variability within the deposit and the effect of this variability on Zn recovery. Recovery is adversely affected by the presence of high barite and other deleterious minerals and ore textures. The existing long term resource model was constructed by independently kriging the four main variables.

Improved multivariate modeling of the different elements and ore types should improve the reliability of the long-term resource model and therefore the prediction of Zn recovery.

5.1 Background

The Red Dog Main Pit consists of three geological plates: Upper, Median and Lower. There are a total of 31 geology codes, of which only eight will be modeled. These eight geological rock types correspond to four different ore type units in two separate plates.

The existing grade models were kriged at a 25ft × 25ft × 25ft ($(25\text{ft})^3$) resolution. For this case study, the geostatistical models will be simulated at 12.5ft × 12.5ft × 12.5ft ($(12.5\text{ft})^3$) resolution, and will later be upscaled to $(25\text{ft})^3$ for comparison purposes. There are some good reasons to model at a finer scale than will be required later. Firstly, the 12.5ft composite data are a good compromise between retaining some of the variability of the smaller drillhole sample data and the faster simulation of larger, and hence fewer blocks. Secondly, the simulation is essentially a "point"-scale simulation; current implementations do not explicitly account for volume-variance relations. Thus, simulating at a finer resolution and then averaging

Direction	Minimum (ft)	Maximum (ft)	Number of Cells	Size of Cells (ft)
Easting	585000	589500	360	12.5
Northing	141500	146000	360	12.5
Elevation	800	950	12	12.5

Table 5.1: Red Dog model coordinate limits.

to larger blocks will show the variability of the block grades more accurately.

Six benches were modeled to allow for model reconciliation with blast hole samples. Table 5.1 lists the coordinate limits of the conditional simulation model. These limits essentially cover the entire areal extents of the Main pit and the vertical extents of the six benches of interest. This model consisted of a total of 1,555,200 cells.

The simulations were constructed on a by rock type basis, and all figures shown correspond to one particular rock type. Once all rock types were simulated, the realizations were merged and all global comparisons consisted of all rock types taken together.

5.2 Available Data

Three types of data were made available by Teck Cominco: drillhole data, composited drillhole data and blasthole data. Multivariate geostatistical modeling considered the 12.5ft composites, while the blasthole data were used to test the predictive ability of the resulting models.

There were a total of 9847 12.5ft composites available for the eight rock types of interest. The term drill hole (DH) refers to the 12.5ft composites. DH data are at a nominal 100ft × 100ft spacing.

For these same rock types, there were 58566 blast hole (BH) data available for model validation. BH data are more closely spaced than DH data at 10ft × 12ft spacing. Figure 5.1 shows the projection of the available data onto a plan and a section view of both data types separately. Note that in the plan and section for BH data, the data density is high, and the distance between the BH samples is very small relative to the size of the field.

A geology model at $(25\text{ft})^3$ resolution was also available. For consistency with the simulation models, the $(25\text{ft})^3$ geology model was reformatted into a $(12.5\text{ft})^3$ model.

5.3 Multivariate Geostatistical Simulation

Conditional simulations were performed for seven variables: Zn, Pb, Fe, Ba, sPb (soluble Pb), Ag, and TOC (total organic content). These seven variables were modeled for each rock type, using Gaussian simulation with stepwise conditionally transformed variables. The main steps of the simulation are:

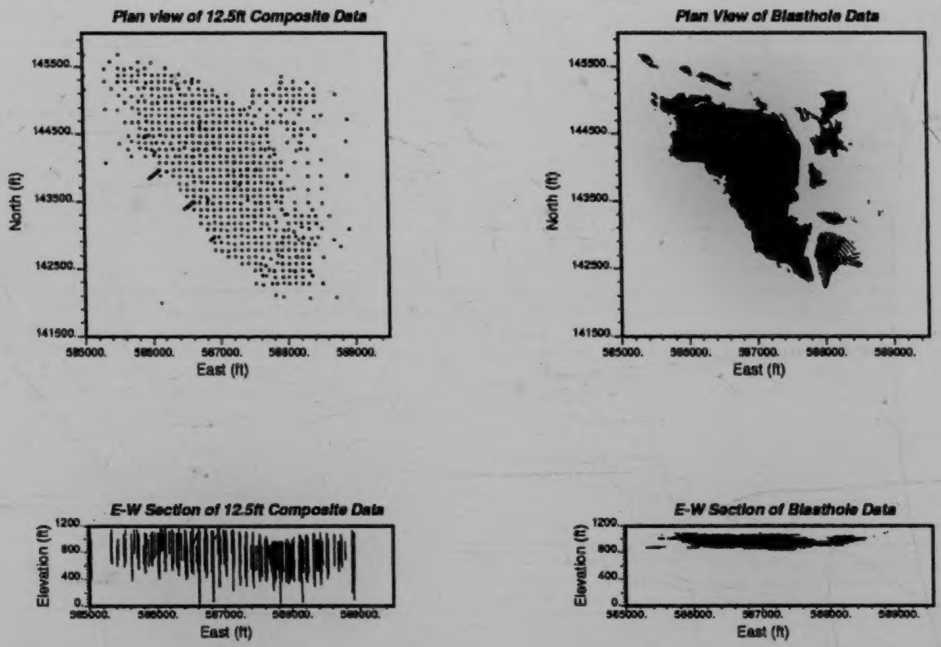


Figure 5.1: All available data projected onto a horizontal plan (plan view (top)) and onto a vertical plane (E-W sectional view (bottom)): 12.5ft composites (left) and blasthole data (right).

Transform No.	Transform Order		
	First Variable	Second Variable	Third Variable
1	Zn	Pb	Fe
2	Zn	Fe	Ba
3	Zn	Pb	sPb
4	Zn	Pb	Ag
5	Zn	Fe	TOC

Table 5.2: Transformation ordering for stepwise conditional transformation.

1. Transform data in a stepwise conditional manner to obtain independent Gaussian variables (Chapter 3).
2. Calculate and model the directional variograms for each of the transformed variables within each rock type (Section 2.1.3).
3. Simulate transformed variables via sequential Gaussian simulation (Section 2.1.7).
4. Back transform simulated values to original units (Chapter 3).

Once all variables within all rock types were modeled, all block models were merged to form multiple realizations of the study area for uncertainty assessment and post-processing. All simulation related tasks were performed using GSLIB [21] and related GSLIB-compatible tools.

The need to model seven variables with only 3000 composites for any one rock type (that is, the rock type with the most data contained only 3000 samples) poses a problem in practice . The multivariate stepwise conditional transform would require 10^7 composites in order to have a minimum of 10 data per probability class. This is impractical. A *nested* application of the stepwise conditional transformation is proposed to overcome this problem. Accounting for a lower-dimensional multivariate distribution was considered. Inference of a trivariate distribution would require approximately 10^3 or 1000 data to define the conditional distributions with a minimum of 10 data. This is more reasonable given the number of composites available.

Recall from Chapter 4 that the transformation ordering for the stepwise conditional transform will affect the reproduction of the variogram from simulation. Thus, the most important variable or the most continuous variable should be chosen as the primary variable (Section 4.4). For Red Dog, Zn is the most important variable, and so all others will be conditioned to it. To account for the other six variables, the following sets of transformations were proposed:

The transformation order reflects the significance Teck Cominco staff attribute to each variable. Zn is considered to be the most important, and so all other variables are transformed conditional to Zn. In all cases, Fe or Pb act as secondary variables, and all remaining variables are then transformed conditional to either Zn and Pb or Zn and Fe.

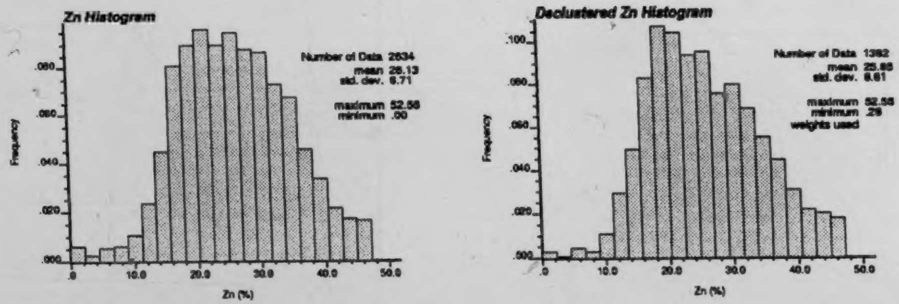


Figure 5.2: Comparison of equally weighted Zn distribution and representative Zn distribution.

Data Declustering. An important aspect of geostatistical simulation is to assemble representative distributions for each variable. Given the multivariate nature of this dataset and the intended application of a multivariate transformation technique, declustering must be consistent between all variables. This consistency involves respecting the multivariate relations and the manner in which simulation will account for them. Multivariate dependency between all seven variables is a direct consequence of the transformation order that will be imposed (see Table 5.2).

The representative distribution of Zn must be established by a declustering procedure using the data configuration and the volume of a particular rock type. For this purpose, kriging within a rock type was performed; the kriging weights given to each data were accumulated, and these weights were then used as the declustering weights. This approach not only respects the rock type being populated (much like nearest neighbour declustering), but it also respects the spatial variability of the data and hence their area of influence within this rock type.

Declustering of the secondary variables (say Pb) must respect the bivariate relations since these will be transformed conditional to Zn. The stepwise conditional transform considers only those secondary data where primary data is available (that is, at locations where there are both Zn and Pb data). As a result, the distribution that will be reproduced in the back transformation is the isotopically sampled values of Pb with Zn. For this reason, a bivariate calibration of the Pb distribution was performed using both the representative distribution of Zn and the relationship between Zn and Pb. For all tertiary variables, the same rationale was applicable, and the representative histograms for Fe through TOC were determined using the representative histograms for the two dependent variables plus the trivariate calibration data.

Figure 5.2 shows the comparison of the histogram of the Zn data within one rock type and the corresponding representative histogram of Zn grade. Note there are 2634 data for the equal weighted histogram on the right. As discussed above, not all of these are used to assemble the representative histogram of Zn; in fact, only about half the total data are used to decluster. Declustering with the kriging weights was applied using the rock type model to assemble the representative Zn histogram.

Only minor changes were apparent. There was no change in the data values, they

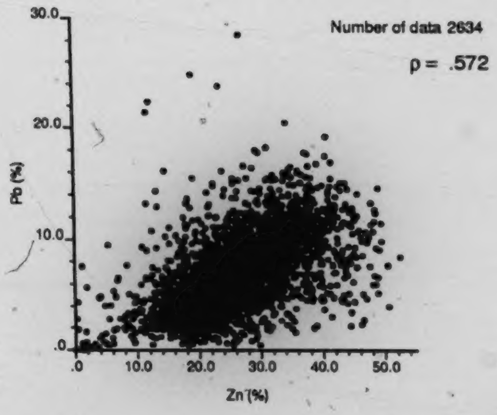


Figure 5.3: Calibration crossplot of Pb given Zn.

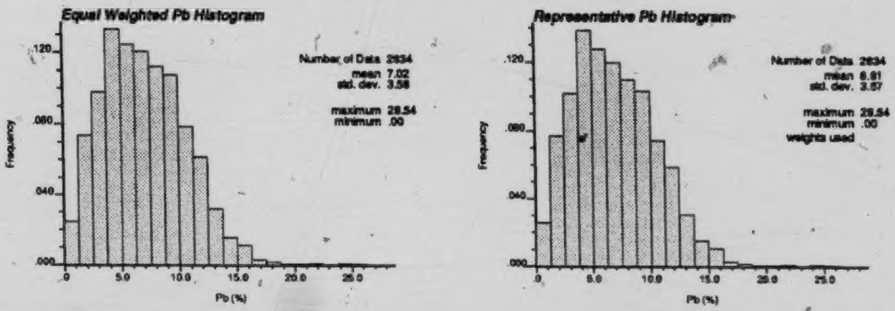


Figure 5.4: Comparison of equally weighted Pb distribution and representative Pb distribution.

were only weighted differently. The mean and standard deviation have decreased slightly. The scatterplot of Pb given Zn was used to calibrate the marginal distribution of the Pb to give the representative Pb distributions (see Figure 5.3). Figure 5.4 shows a comparison between the equal weighted histogram of all Pb data (using 2634 Pb data) and the representative Pb histogram. Given the positive correlation between Pb and Zn and the slight decrease in the mean of Zn, it was expected that the representative histogram of Pb should have a slightly lower mean than the 7.02% reported on the equal weighted histogram; and it has indeed decreased to 6.91 %.

Stepwise Conditional Transform of Red Dog Data. Figure 5.5 shows the scatterplots of the variables resulting from the first transform sequence of Zn, Pb and Fe (see Table 5.2).

The transformed variables are independent and multiGaussian, which translates to a circular shape in the crossplot. From Figure 5.5, the crossplot between the first two variables (Zn and Pb) appears approximately circular. Crossplots with the third variable (Fe, in this case) show some banding; however this is simply a visual

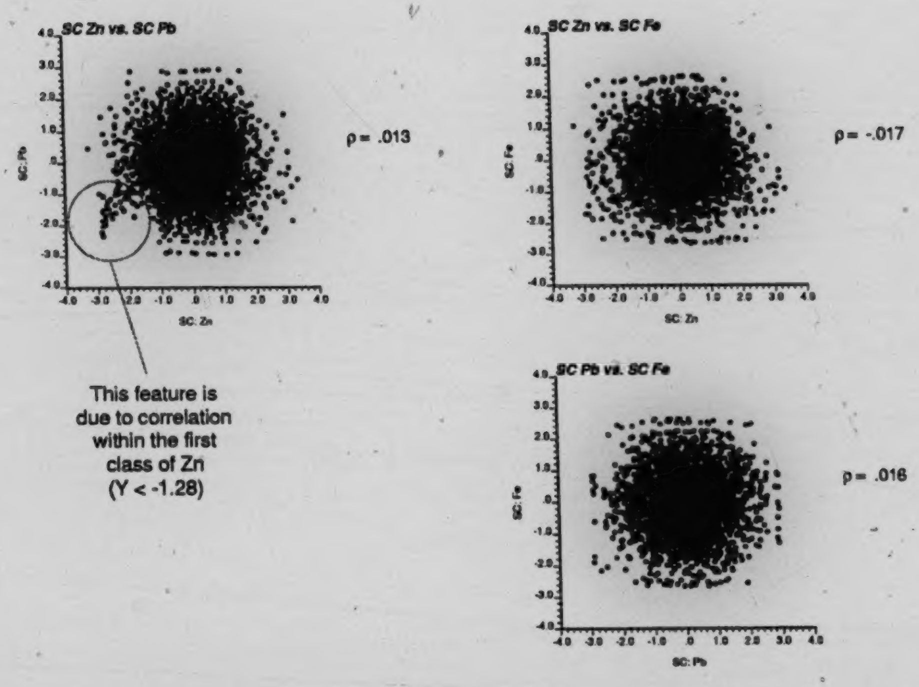


Figure 5.5: Crossplot between stepwise conditionally transformed variables for Zn, Pb and Fe. Zn was transformed first, then Pb was transformed conditional to Zn, and finally Fe was transformed conditional to both Zn and Pb.

artefact of having many classes and consequently fewer data within each class (see Section 4.2).

Variogram Analysis. Spatial statistics can now be calculated and modeled for the transformed data. The variography was determined for the transformed variables.

Figure 5.6 shows an example of variogram maps, experimental variograms and the variogram model for Zn. Both the variogram model and the experimental variogram points are shown in the chosen principal directions. Overall, the experimental variograms are fairly stable and the corresponding models fit the experimental points well.

Variogram maps are a common tool used to attain a preliminary impression of continuity directions. The maps are in radial coordinates; at the centre of the map is location $\mathbf{h} = 0$. Basically, the maps are used to visualize large scale continuity directions and distances. These directions and distances are then further refined during the calculation and modeling of experimental variograms.

Simulation. Sequential Gaussian simulation (SGS) was independently performed for each of the seven transformed variables on a by rock type basis: Zn, Pb, Fe, Ba, sPb, Ag and TOC. A total of 40 realizations were generated for each variable within each rock type. For greater computational efficiency, only those blocks belonging to the specific rock type were simulated (as controlled by the geology model).

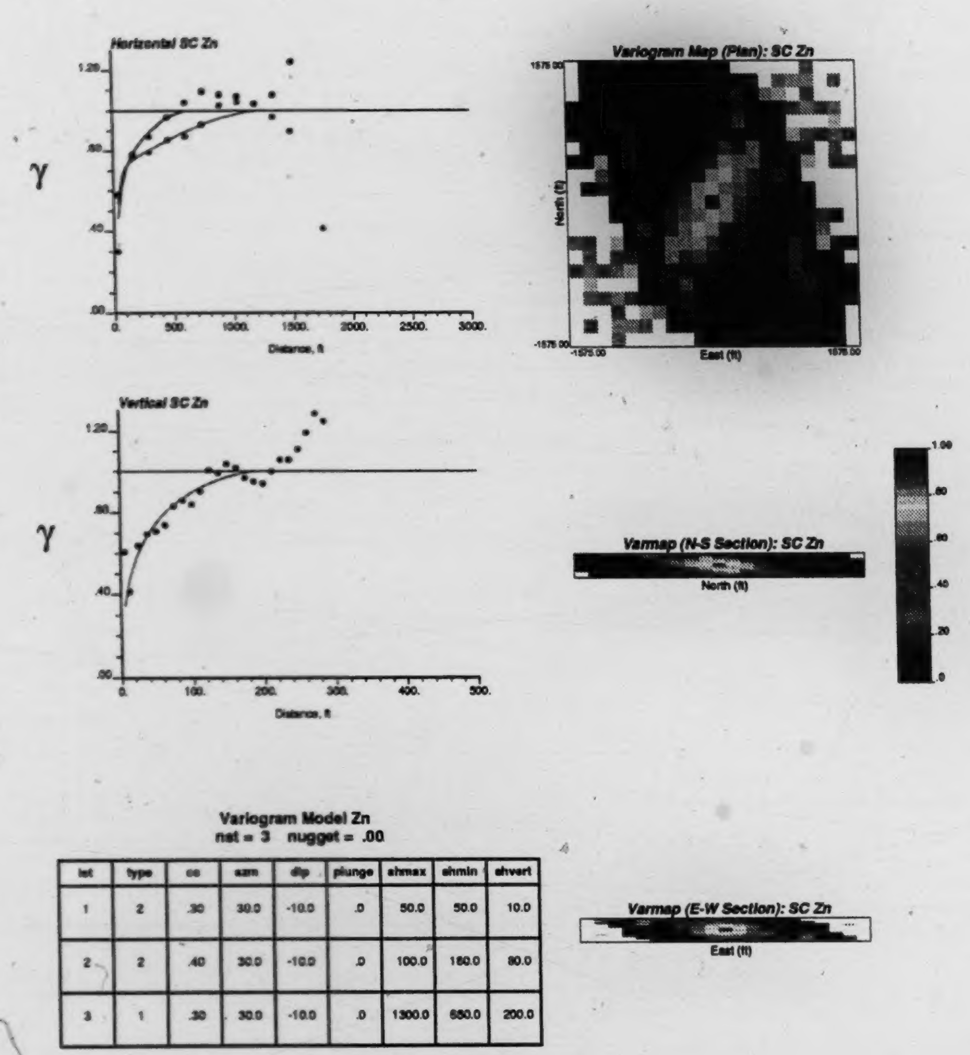
Back Transformation. The simulation results must be back transformed to the original units of the data. Similar to the forward transformation that relied on conditioning one variable to another, the back transformation for each simulated realization must be performed in a conditional fashion. For example, the back transform of Fe will be conditional to the simulated values for Zn and Pb. The key to back transformation is to maintain consistency with the forward transformation.

5.3.1 Validation of Simulation Models

A number of basic checks must be performed prior to using these models for decision making.

An important validation is the reproduction of the input data and the variogram. Statistical fluctuations are inherent in stochastic simulation; however, these fluctuations should be reasonable and unbiased. Simulation produces simulated values that are approximately standard normal in *expected* value. For any one realization, minor fluctuations from a zero mean and unit variance are expected; however, when these values are back transformed to original units a slight shift of the mean in normal space may translate to a more significant shift of the mean in original units. Similarly, the combined fluctuation of the mean and variance in normal space may translate to more noticeable shifts in original space. This is particularly true for skewed distributions, which is the case for most variables in the Red Dog data.

Deviations from the limit standard normal distribution could be due to a number of factors. Firstly, the algorithms employed are based on an assumption of



Variogram Model Zn
nst = 3 nugget = .00

ist	type	cc	azm	dip	plunge	ahmax	ahmin	ahvert
1	2	.30	30.0	-10.0	.0	50.0	50.0	10.0
2	2	.40	30.0	-10.0	.0	100.0	180.0	80.0
3	1	.30	30.0	-10.0	.0	1300.0	650.0	200.0

Figure 5.6: Variogram model results for Zn. Horizontal variograms (top left) show maximum continuity direction (N30E) in grey and minimum continuity direction in black. Vertical variogram (middle left) and corresponding variogram model parameters (bottom left). Variogram maps are shown on the right: plan view (top), N-S section looking west (middle), and E-W section looking north (bottom).

stationarity. Non-stationary data can lead to shifts in the mean and/or variance of simulated values in normal space. Secondly, Gaussian simulation techniques assume the data are multiGaussian in a spatial context. There are no techniques to ensure multiGaussianity in the spatial domain. To mitigate the effects of fluctuations in normal space and its translation to original space of the data, a standard transform is applied to the simulated values to ensure reproduction of the histogram and its corresponding summary statistics [41]:

$$z_2(\mathbf{u}) = z_0(\mathbf{u}) + \lambda(\mathbf{u})[z_1(\mathbf{u}) - z_0(\mathbf{u})] \tag{5.1}$$

where

$F_0(z)$ = cdf of the simulated values
$F_1(z)$ = target cdf
$z_0(\mathbf{u})$ = set of originally simulated values, $\mathbf{u} \in A$
$z_1(\mathbf{u})$ = corrected value based on quantile transformation alone
= $F_1^{-1}(F_0(z_0(\mathbf{u})))$
$z_2(\mathbf{u})$ = corrected value based on quantile transform and kriging variance
$\lambda(\mathbf{u})$ = correction factor defined as $[\sigma_K(\mathbf{u})/\sigma_{max}]^{\omega}$
$\sigma_K(\mathbf{u})$ = kriging variance at location $\mathbf{u}$
$\sigma_{max})$ = $\max\{\sigma_K(\mathbf{u}), \mathbf{u} \in A\}$
ω = correction level parameter, must be > 0

Use of the kriging variance in Equation 5.1 ensures that the values at data locations are reproduced. The transform is applied over individual realizations to ensure reproduction of the global histogram for each realization. Alternatively, "sets" of realizations can also be transformed and data would still be honoured; however, this does not guarantee that the global histogram per realization is reproduced.

Data Reproduction. The goal is to verify that the corresponding simulated values reproduce the assigned composite values. For each variable, a crossplot showing the DH values and their corresponding simulated values demonstrates whether input data were reproduced within the numerical precision of the storage and the transformation table being used. Figure 5.7 shows an example of the crossplot for Zn. Over the multiple rock types and multiple variables, the composite data were reproduced exactly in almost all cases. Deviations from the true value were a result of the numerical precision reported in the transformation table. In general, minor fluctuations at the high and low ends of the units were attributed to the number of quantiles reported in the transformation table. Overall, DH data were reproduced at their respective locations from the simulation models.

Histogram Reproduction

Another important check is the histogram of the simulated values after back transformation. These distributions should be similar to the representative histograms, with comparable statistics. Figure 5.8 shows an example of one such comparison for

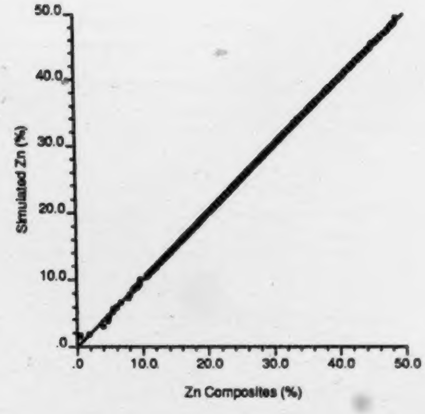


Figure 5.7: Composites data reproduction for Zn.

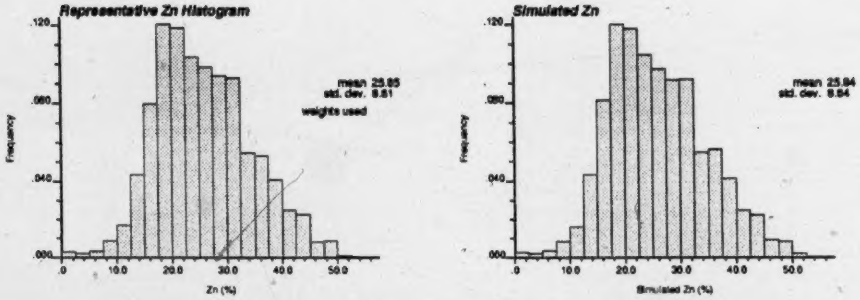


Figure 5.8: Histogram reproduction for Zn: representative Zn histogram (left) compared to simulated Zn histogram (right).

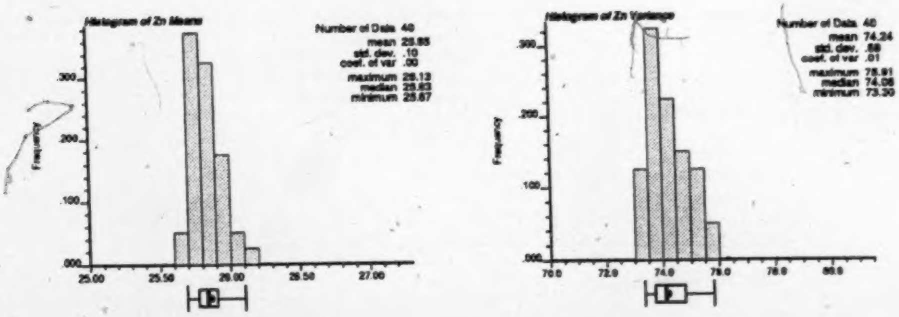


Figure 5.9: Reproduction of summary statistics for Zn: histogram of means (left) and variances (right) from multiple simulated realizations. Box plots on x-axis shows the 95% probability interval (outside lines), 50% probability interval (box) and the median (vertical bold line inside box). The dot indicates the mean value of the summary statistic from the declustered distributions.

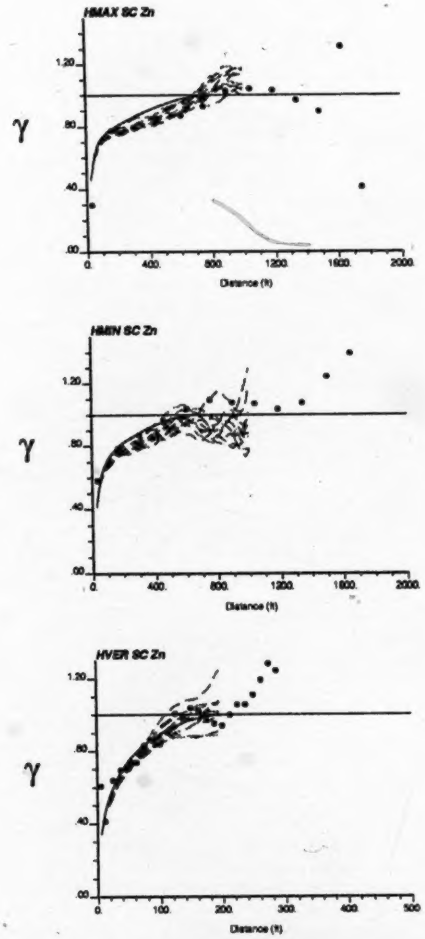
Zn. Overall, the histograms were reproduced within reasonable statistical fluctuations in the summary statistics by construction (see Equation 5.1).

Over the 40 realizations (or the ensemble), the summary statistics can be checked for reproduction. For each realization, there is a mean and a variance associated to the resulting global distribution. A histogram of the mean from all realizations will show if the mean is reproduced. Similarly, a histogram of the variance will show if the global variance is reproduced. Figure 5.9 shows these two histograms for Zn. The mean of the representative distribution was reproduced; differences in magnitude lie in the second decimal place. As well, the distribution of variances showed only minor differences in magnitude. This reproduction was a consequence of the transform applied (Equation 5.1).

Variogram Reproduction. After verification that the first order statistics were satisfactorily reproduced, the next check involved the variogram. It is important to note that this check was performed in normal or transformed space (prior to back transformation), since only the normal scores variogram will be reproduced. Figure 5.10 shows the results of performing this check for Zn. The variogram directions calculated from the simulated models correspond to the standard North-South (N-S), East-West (E-W), and vertical directions. These variograms are shown as grey dashed lines. The variogram models were calculated in these same three directions (although principal directions may differ) and are shown as black solid lines. The experimental variogram points calculated from the 12.5ft composite data are also shown, and correspond to the black dots on these figures.

Reproduction of Multivariate Features. The multivariate relations are important and must be checked. Teck Cominco was also interested to see how their existing models "performed" for the same type of check.

In order to allow for direct comparisons between a simulation and the existing



Variogram Model Zn
nst = 3 nugget = .00

ist	type	cc	azm	dip	plunge	ahmax	ahmin	ahvert
1	2	.30	30.0	-10.0	.0	50.0	50.0	10.0
2	2	.40	30.0	-10.0	.0	100.0	180.0	80.0
3	1	.30	30.0	-10.0	.0	1300.0	650.0	200.0

Figure 5.10: Variogram reproduction for Zn: horizontal maximum direction (top), horizontal minimum direction (second), vertical direction (third), and table listing variogram model. Black solid line represents the variogram model, and dashed lines represent the variogram of the simulated models.

long term model, the simulations must first be upscaled to the same volume as the existing model. This required upscaling from (12.5ft)^3 models to (25ft)^3 models. This upscaling is discussed in more detail below in Section 5.4.

Figure 5.11 shows a comparison of the crossplot reproduction from simulation to those crossplots from the 25ft composites and the existing long term model. In general, the simulated realizations reproduce the trivariate relations with comparable variability to the 25ft composites; the corresponding plots from the existing long term model shows similar bivariate relations but with noticeably reduced variability.

Recall that the relations between Zn and Ba was the most important for Zn recovery. Comparing the Zn-Ba crossplot from all three sources (composites, simulation, and long term model) shows the existing model reproduced neither the bivariate relations nor the inherent variability of the data. This result and its potential impact on production supports the use of multivariate geostatistics in model construction.

5.3.2 General Comments on Conditional Simulation Models

Once all simulated models were generated and validated on a by rock type basis, a single realization for each variable was obtained by merging the simulated properties from each rock type. Figure 5.12 shows a few of the simulated realizations for Zn at the 12.5ft grid resolution.

The modeling methodology implemented in this project was quite complex. Conventional approaches are sufficient for straightforward problems; however, for the complexity of the Red Dog data, these common approaches are inadequate. The availability of multiple metal grades within multiple rock types warrants some consideration of the relationship between these grades and how these relationships change from one rock type to the next. The approach documented in this section was designed to explicitly address this key issue. Consequently, the resulting models not only reproduce the univariate data and its spatial variability, but taken together, they also honour the multivariate relations between the different metals/minerals within the different rock types.

5.4 Validation with Additional Data

Blasthole (BH) data was intentionally excluded from the input data used for model construction. The idea was to assess the predictive ability of the conditional simulation models using the BH data. In particular, four variables will be compared: Zn, Pb, Fe and Ba.

Comparing BH and DH Data. The BH data were paired with the closest DH data within a tolerance of 50ft. A total of 9846 DH composites were checked against 58560 BH data. Figure 5.13 shows the results of pairing up the BH with the DH data. Significant banding was expected as a result of pairing the DH data with multiple BH data.

This nearest-neighbour pairing of the BH and DH data presented the lower bound on the expected correlation. Simulated values were generated using kriging

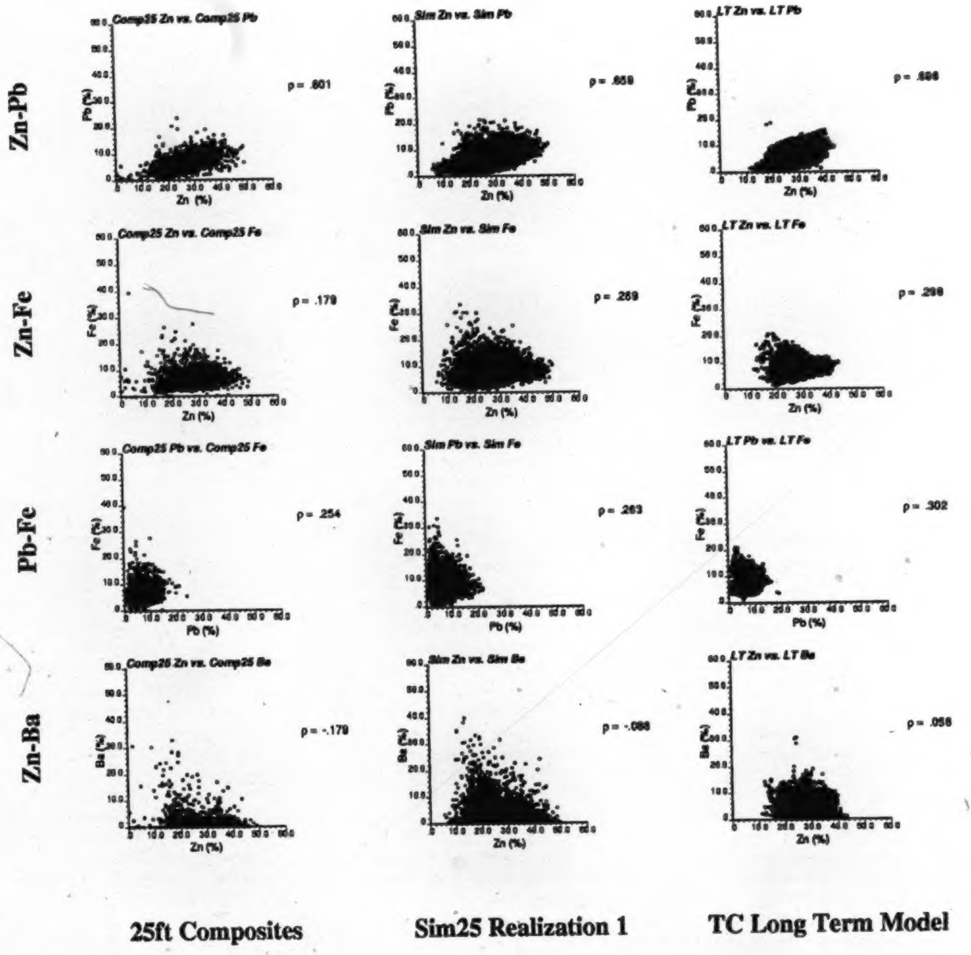
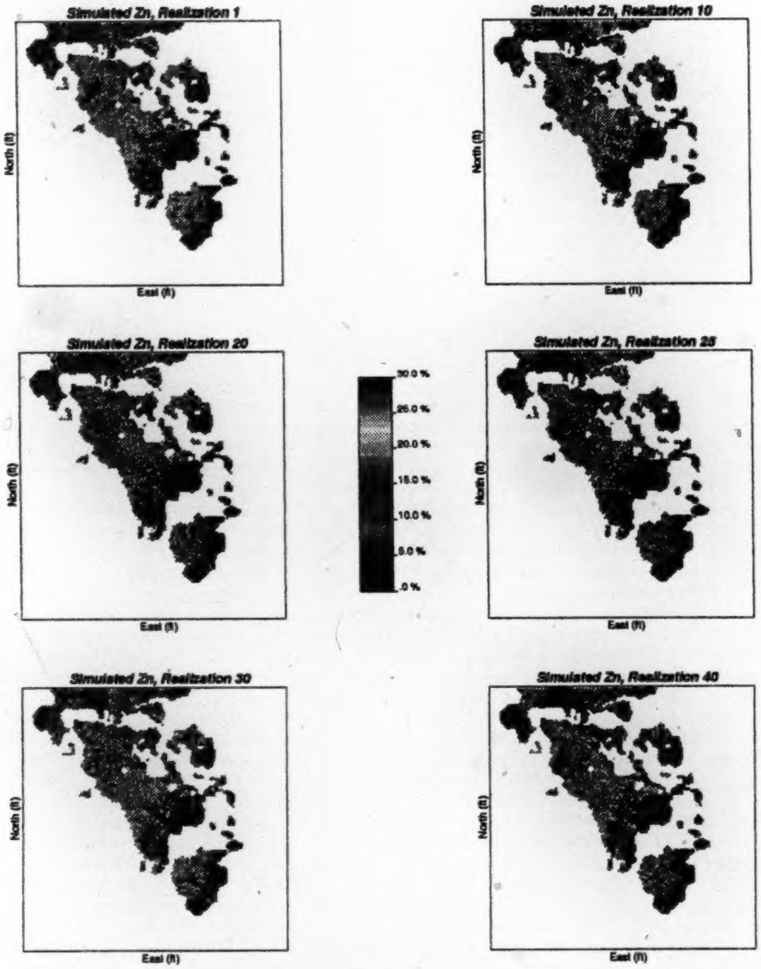


Figure 5.11: Comparison of multivariate features reproduction for Zn-Pb (top row), Zn-Fe (second row), Pb-Fe (third row), and Zn-Ba (bottom row). Crossplots using the 25ft composites are shown on the left column, from the upscaled simulations are shown in the middle column and those from the existing long term model are shown in the right column.



Simulated Zn, Realization 1

Simulated Zn, Realization 10

Simulated Zn, Realization 20

Simulated Zn, Realization 25

Simulated Zn, Realization 30

Simulated Zn, Realization 40

North (ft)
East (ft)
30.0 %
25.0 %
20.0 %
15.0 %
10.0 %
5.0 %
.0 %

Figure 5.12: Simulated realizations of Zn at 12.5ft grid resolution. This section spans elevations 862.5 to 875ft.

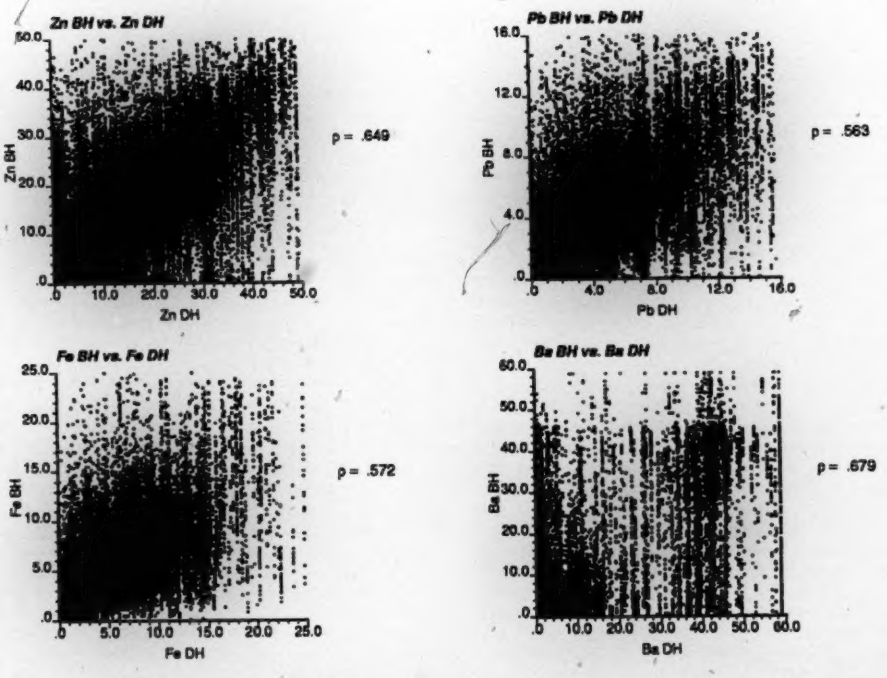


Figure 5.13: Crossplot of BH data against nearest neighbour DH data for Zn, Pb, Fe and Ba over all eight rock types. Significant banding is a result of pairing DH data to multiple BH data.

which accounts for data redundancy, closeness, and the surrounding data values. Having considered the spatial correlations in addition to the data values, the simulated values should be better correlated to BH data. Of course, the variogram used in simulation will have an impact on how much better the expected correlation should be. For instance, a higher nugget effect would reduce the correlation between the simulated value and the BH data. Alternatively, a variogram model with a long range and low nugget effect should increase the correlation between the simulated value and the BH data.

Comparing BH and Conditional Simulation Models. Prior to any type of comparative studies, the $(12.5\text{ft})^3$ models must first be upscaled to a resolution comparable to the BH data, which are at 10 x 12ft areal spacing with a length of 25ft. The models were upscaled to 12.5 x 12.5 x 25ft for consistency with BH data. The block averaging method was a weighted average based on specific gravity equations, provided by Teck Cominco. Figure 5.14 shows a few of the realizations of Zn at this resolution. As a result of upscaling in the vertical direction, these realizations appear slightly smoother than those in Figure 5.12.

Given that there were 40 realizations for the conditional simulation models for each variable, there was the issue of which realization should be compared against

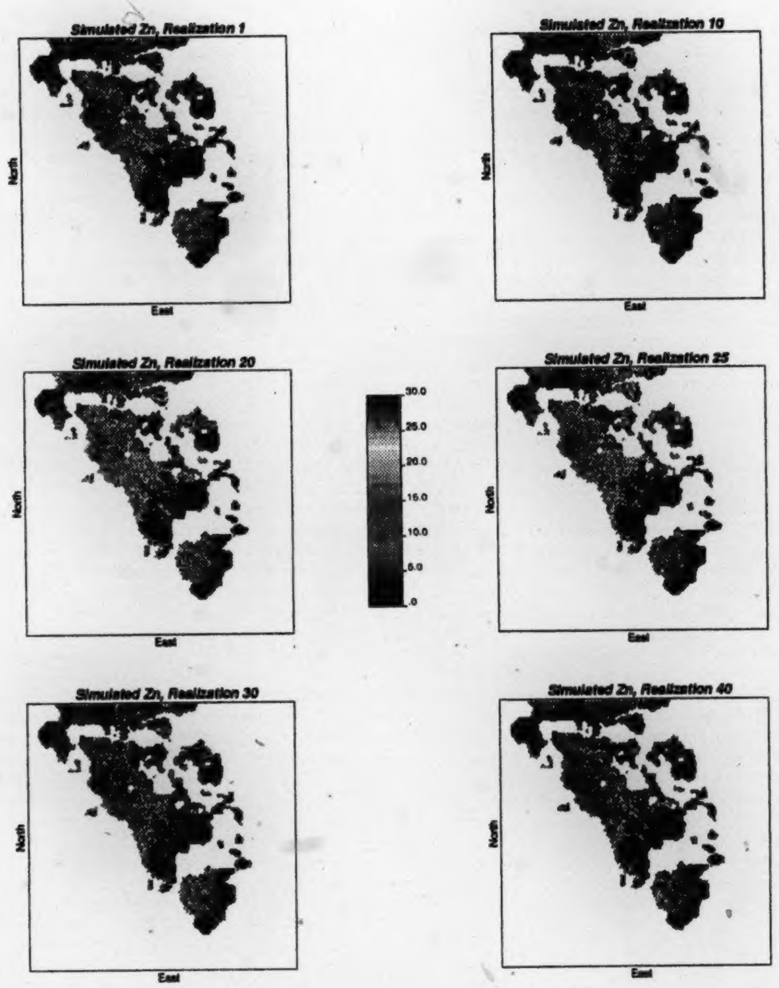

Simulated Zn, Realization 1
Simulated Zn, Realization 10
Simulated Zn, Realization 20
Simulated Zn, Realization 25
Simulated Zn, Realization 30
Simulated Zn, Realization 40
North
East
30.0
25.0
20.0
15.0
10.0
5.0
.0

Figure 5.14: Simulated realizations of Zn at 12.5 x 12.5 x 25ft grid resolution, consistent with BH data. The section shown spans elevations 850 to 875ft.

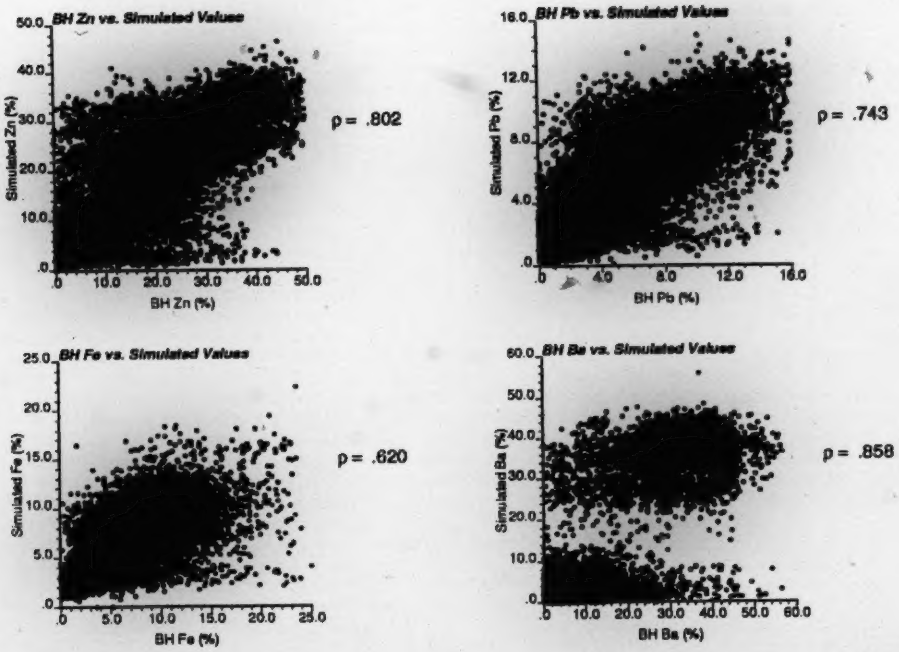


Figure 5.15: Crossplots of BH data and E-type estimate from Conditional Simulation Models.

the BH data. Rather than choose any arbitrary realization (since all are equally likely to be chosen), the E-type estimate was compared. The E-type estimate refers to the expected value at each location calculated based on the local distribution constructed using the 40 realizations. This is similar to a kriged model, since it is a model of expected values. The effect of the variability inherent in any one realization will be mitigated by choosing the E-type estimate.

Figure 5.15 shows the comparison between the BH data and the E-type estimate. The crossplots show fairly strong positive correlations between the model and the BH data, ranging from 0.62 to 0.86. Two distinct populations were apparent from the Ba crossplot, which represented the two rock types. A comparison with Figure 5.13 shows that for all four variables, the correlation between the BH data and the simulated values were higher than the lower bound represented by the pairing of BH to DH data.

Determining Expected Correlations. It is possible to determine the expected correlation coefficient between the simulation models and the BH data, given the known DH and BH spacing, and the variogram model used to construct the simulations. Figure 5.16 shows a schematic illustration of the relationship between the BH data spacing, variogram model and the expected correlation. Use of this approach assumed that the BH data and the DH data have the same variogram model, that is, the spatial variability of the BH data was the same as that of the DH data. As

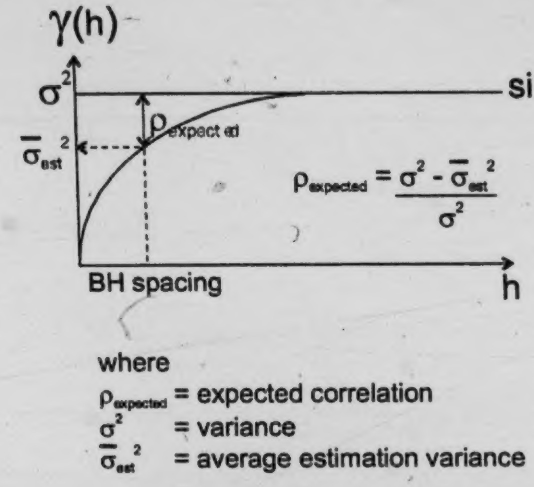

$\gamma(h)$
σ^2
sill
$\overline{\sigma}_{est}^{\,2}$
$\rho_{expected}$
$\rho_{expected} = \frac{\sigma^2 - \overline{\sigma}_{est}^{\,2}}{\sigma^2}$
BH spacing
h
where
$\rho_{expected}$ = expected correlation
σ^2 = variance
$\overline{\sigma}_{est}^{\,2}$ = average estimation variance

Figure 5.16: Schematic illustration of relation between expected correlation and estimation variance at BH spacing. This is calculated for multiple BH locations, and then averaged to obtain the average estimation variance.

well, the BH data were assumed to have similar statistical properties as the DH data.

From Figure 5.16, the key to determining the expected correlation was to first determine the average value of the estimation variance at the BH data spacing. Recall that the DH data are at 100 x 100ft spacing while the BH data are at 10 x 12ft spacing. As a result, BH data are interspersed between the DH data, and so the estimation variance must be calculated at each BH location and then averaged within the 100 x 100ft block. For the purposes of approximating this value, the model's 12.5 x 12.5ft horizontal spacing was sufficient for substituting the actual 10 x 12ft BH spacing.

Since the variogram models were constructed for the normal scores, the variance is 1.0. The expected correlation is given by:

$$\rho_{expected} = 1.0 - \overline{\sigma^2_{est}}$$

Figure 5.17 shows the configuration of the DH data and the block model spacing. For simplicity, the BH data were assumed to be centered about a DH sample on an approximate 8 x 8 grid with 12.5 x 12.5ft blocks. Kriging can be performed on this data configuration with an appropriate variogram model to determine the estimation variance for each of the 64 blocks. Averaging these values yields the average estimation variance of the BH data surrounding a DH data. For example, kriging was performed using the Zn variogram for the configuration shown in Figure 5.17. Specific DH values at the 100 x 100ft spacing were not important since we were only concerned with the estimation variance (recall that the estimation variance is not dependent on the data value, but rather the data locations (Section 2.1.6).

Figure 5.18 shows the histogram of estimation variances for the 64 blocks ob-

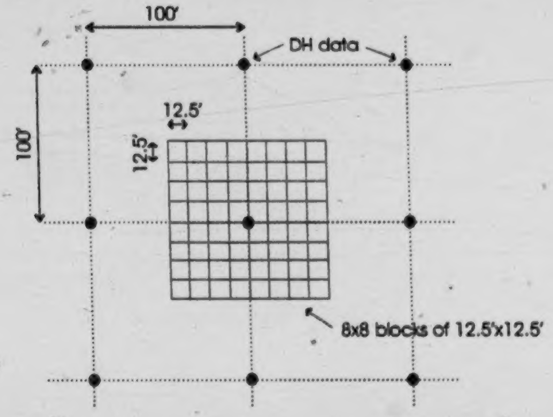


Figure 5.17: Configuration for determining average estimation variance. DH data are at 100 x 100ft spacing, BH data are at 10 x 12ft spacing. Set up an 8 x 8 grid centred about a DH data, with block sizes of 12.5 x 12.5 x 25ft. Perform kriging to determine the estimation variance at each block, and then average these to get the mean estimation variance.

Variable	$\bar{\sigma}^2_{est}$	$\rho_{expected}$	$\rho_{BH-E-type}$
Zn	0.658	0.342	0.522
Pb	0.647	0.353	0.695
Fe	0.577	0.423	0.648
Ba	0.834	0.166	0.375

Table 5.3: Summary table for determining expected correlation coefficient for one rock type: average estimation variance from kriging (second column), and expected correlation coefficient (third column), and correlation between BH and E-type estimate from simulation models (fourth column). For each variable, the actual correlation exceeds the expected correlation.

tained from kriging and the crossplot of the BH data to the E-type estimate. From the histogram, the mean estimation variance is 0.658. From the above equation, the expected correlation is (1.0-0.658) or 0.342. The crossplot of the E-type estimate and the closest BH data shows a correlation of 0.522, which exceeds the expected correlation.

Table 5.3 summarizes the results of calculating the average estimation variance and the corresponding expected correlation for each variable within one rock type, and compares this with the correlation between the BH data and the E-type estimate of the simulation models.

A comparison of the last two columns in Table 5.3 shows that for all four variables, the correlation between the E-type estimates of the models was higher than the expected correlation determined from the variogram models. It was understandable that these correlations may appear unreasonably low; however, the correlations

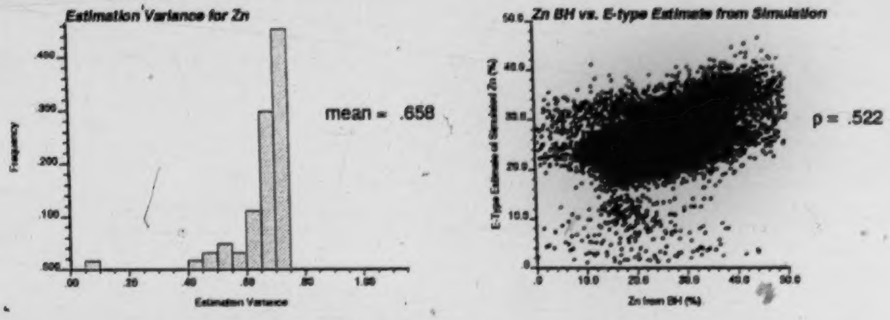


Figure 5.18: Calculation of expected correlation for Zn within one rock type: Histogram of estimation variance from kriging for 8 x 8 grid centered about a DH sample (left), and crossplot of BH data and E-type estimate (right).

Variable	ρ_{BH-LT}
Zn	0.417
Pb	0.439
Fe	0.559
Ba	0.298

Table 5.4: Correlation between BH data and Long Term Model values.

obtained from the models were comparable to the expected correlation.

Similar correlations were expected from the long term model, if the comparisons were carried out on a by rock type basis. Table 5.4 summarizes the correlation coefficients from this type of comparison using the long term model. As expected, these numbers were comparable to the fourth column in Table 5.3. Unfortunately, a direct comparison of these two sets of correlation coefficients would be technically incorrect since the expected correlations for the long term model were a function of the variography used to construct the simulation models. The appropriate variogam models for the long term model were not available.

Model Accuracy and Precision. Another interesting measure of model validation was to assess its accuracy and precision. In a statistical context, a model is considered accurate if for a given symmetric probability interval *p*, the fraction of true values falling within the *p* interval is greater than or equal to *p* for all *p* in [0, 1]. For instance, for a probability interval of 50%, an accurate model should have at least 50% of the true values falling within this interval. The term precision refers to how close this fraction of true values is to *p* for all *p* in [0, 1] (Deutsch, 2002). These two closely related terms are neatly characterized by a simple crossplot of the fraction of true values against the corresponding probability interval. A model is both accurate and precise if this crossplot shows a 45 degree line. Figure 5.19 shows the crossplot for the E-type estimate of Zn, with the BH data taken to be the true data. For the 70% probability interval, the fraction of the true values falling within

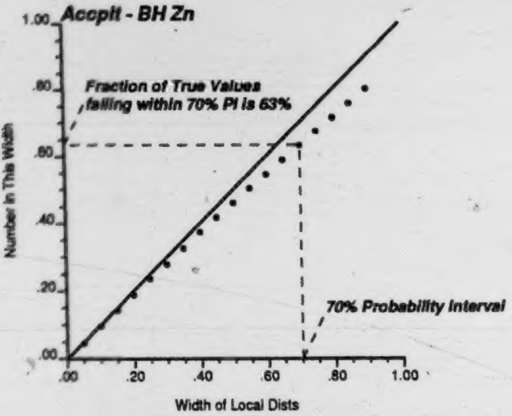


Figure 5.19: Accuracy plot for E-type estimate of Zn data.

this interval is 63%. For this and all other intervals shown in the plot, the match between these two numbers was sufficiently close to indicate that the Zn models were fairly accurate and precise. Figure 5.20 shows the crossplots corresponding to Zn, Pb, Fe and Ba (Zn crossplot is identical to that shown in Figure 5.19). For all four variables, the plots show that the models were satisfactory in their accuracy and precision, with the Fe models as the most accurate and precise.

Scaling up to 25ft blocks. For each of the merged realizations, the $(12.5\text{ft})^3$ grid is scaled up to a $(25\text{ft})^3$ grid to match the resolution of the existing grade and geology models. Similar to the upscaled models for BH comparison, upscaling to $(25\text{ft})^3$ blocks was performed by a weighted average based on specific gravity equations. Figure 5.21 shows a few of the simulated realizations for Zn at the 25ft grid resolution.

Comparison between Simulation to Existing Long-Term Model. Figure 5.22 shows a visual comparison between the E-type estimate from simulation and the existing long term model for bench 850. As expected, both maps were smooth. The E-type estimate should be similar to the kriged results because the E-type is the expected value taken over 40 realizations at each location within the block model and kriging gives the expected value at each location. In contrast, a visual comparison between Figure 5.21 and Figure 5.22 shows that a simulated realization has greater variability than either the long term model (which was kriged) or the E-type estimate. Figure 5.23 shows the crossplot comparison of the E-type estimates from simulation to the corresponding long-term model values. The correlations between the two approaches were high in the case of all four primary variables. The comparison between Ba showed the most differences in the crossplot, with high simulated values paired with some corresponding low long-term model values, and vice versa. Despite this, there was a strong positive correlation. Overall, the high correlations between the two modeling approaches were encouraging statistics that

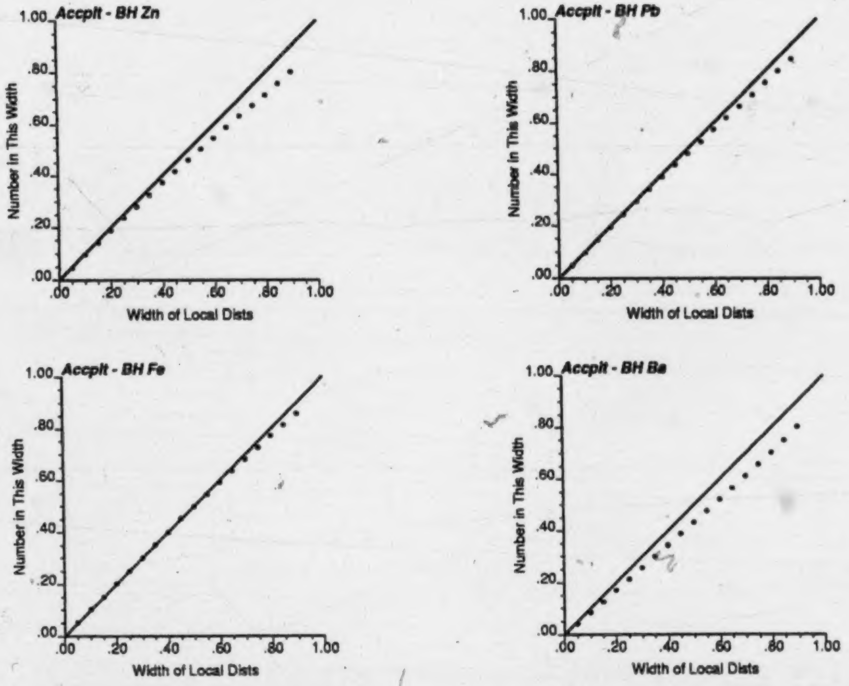


Figure 5.20: Accuracy plot for BH data and E-type estimate of simulation models: Zn (top left), Pb (top right), Fe (bottom left) and Ba (bottom right).

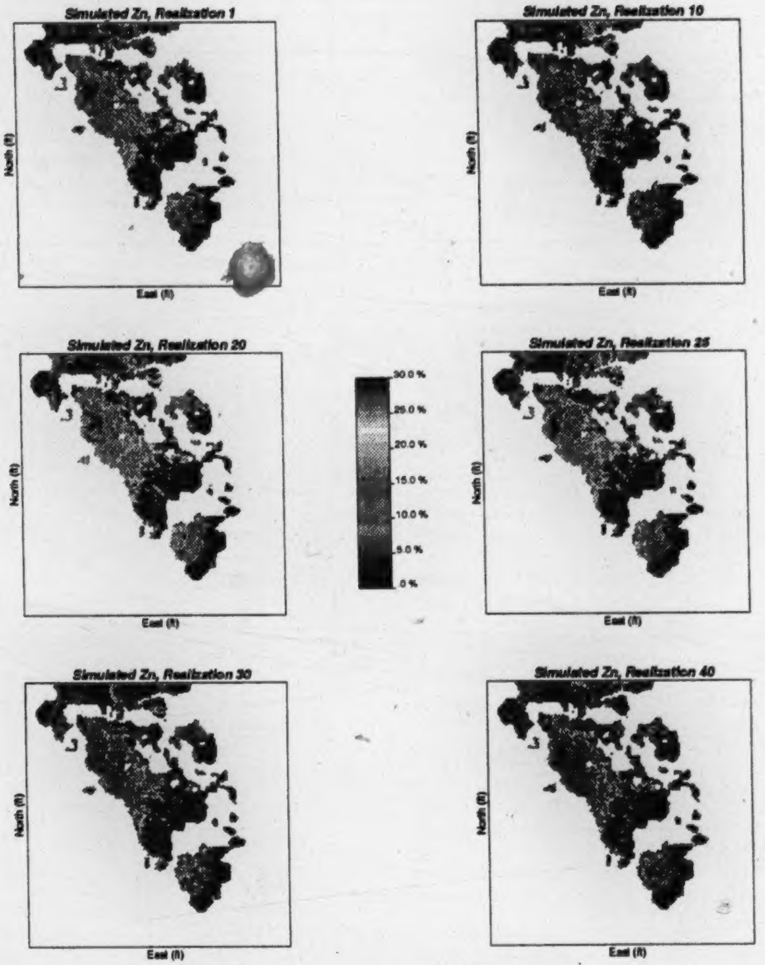

Simulated Zn, Realization 1
Simulated Zn, Realization 10
Simulated Zn, Realization 20
Simulated Zn, Realization 25
Simulated Zn, Realization 30
Simulated Zn, Realization 40
North (ft)
East (ft)
30.0 %
25.0 %
20.0 %
15.0 %
10.0 %
5.0 %
0 %

Figure 5.21: Simulated realizations of Zn at 25ft grid resolution. The section shown corresponds to bench 850 (spanning elevations 850 to 875ft).

Variable	BH-DH	BH-LT Model	BH-E-type	LT Model - E-Type
Zn	0.649	0.803	0.802	0.979
Pb	0.563	0.696	0.743	0.894
Fe	0.572	0.641	0.620	0.871
Ba	0.679	0.856	0.858	0.956

Table 5.5: Summary of correlation coefficients from all comparisons: BH to DH, BH to Long Term (LT) model, BH to E-type estimate, and Long Term model to E-type estimate.

provide validation for both the simulations and the existing models.

Summary of Comparisons

This section described the comparisons between the BH and the DH data, the BH and the E-type estimate of the simulations, the BH and the existing long-term model, and the E-type estimates to the long term model. Table 5.5 gives the summary of the correlation coefficients for these comparisons. From Table 5.5, the worst correlations were given by the BH-DH comparison, which was expected given that the paired BH to DH samples may be separated by distances of up to 50ft. This type of comparison was consistent with a comparison between a simulated or E-type model and the BH data, since model values far away from DH samples are estimates in themselves and also suffer from being far away from real DH data. The comparison between the BH and both the long term model and the E-type estimate from simulation showed very similar correlations, thus indicating that both models have similar predictive abilities.

Given these comparable results, the conditional simulations were considered an improvement over the existing model in that multivariate relations were honoured. The long term model was generated using independent kriging of each variable. Consequently, there was no assurance to honour the multivariate relations between the different metals. For example, the grade of Zn at any one location has no effect on the modeled grade of Pb, Fe or Ba at that same location. In contrast, the simulations were constructed by explicitly accounting for the multivariate relations between the different variables. As a result, the grade of one variable would affect the simulated value of other metals at the same location. Furthermore, the multiple realizations from simulation allow for the assessment of uncertainty on both a local and global scale for decision making.

5.5 Potential Applications of Simulated Models

One of the main benefits of conditional simulation is the ability to assess uncertainty in the model results. There are many ways that the information from multiple realizations can be exploited to yield meaningful results for mine planning and risk assessment. This section discusses a few simple applications such as assessment of

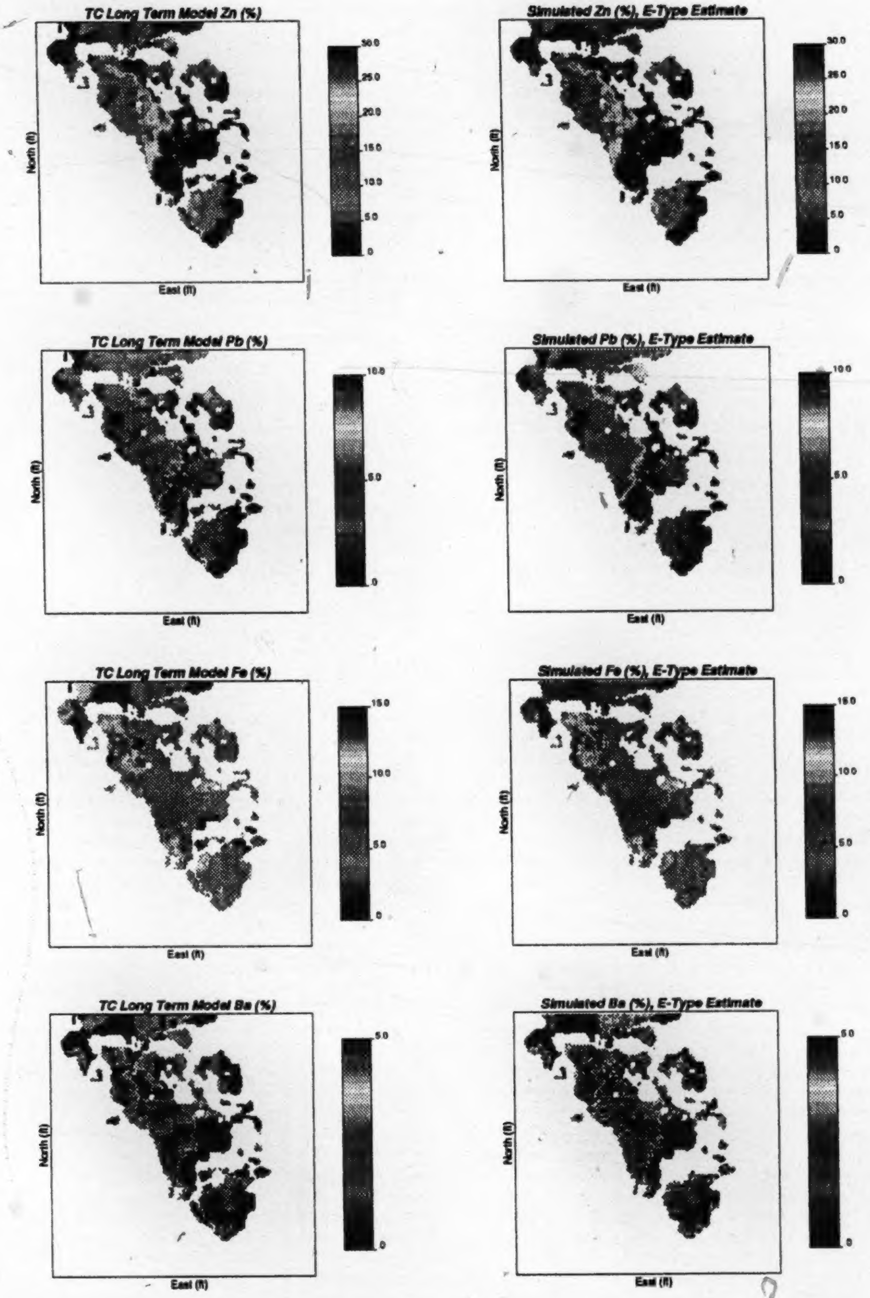

TC Long Term Model Zn (%)
Simulated Zn (%), E-Type Estimate
TC Long Term Model Pb (%)
Simulated Pb (%), E-Type Estimate
TC Long Term Model Fe (%)
Simulated Fe (%), E-Type Estimate
TC Long Term Model Ba (%)
Simulated Ba (%), E-Type Estimate
North (ft)
East (ft)
30.0
25.0
20.0
15.0
10.0
5.0
0

Figure 5.22: Comparison between existing long term model (left) with E-type estimate from simulation (right) at 25ft grid resolution. The section shown corresponds to bench 850 (spanning elevations 850 to 875ft).

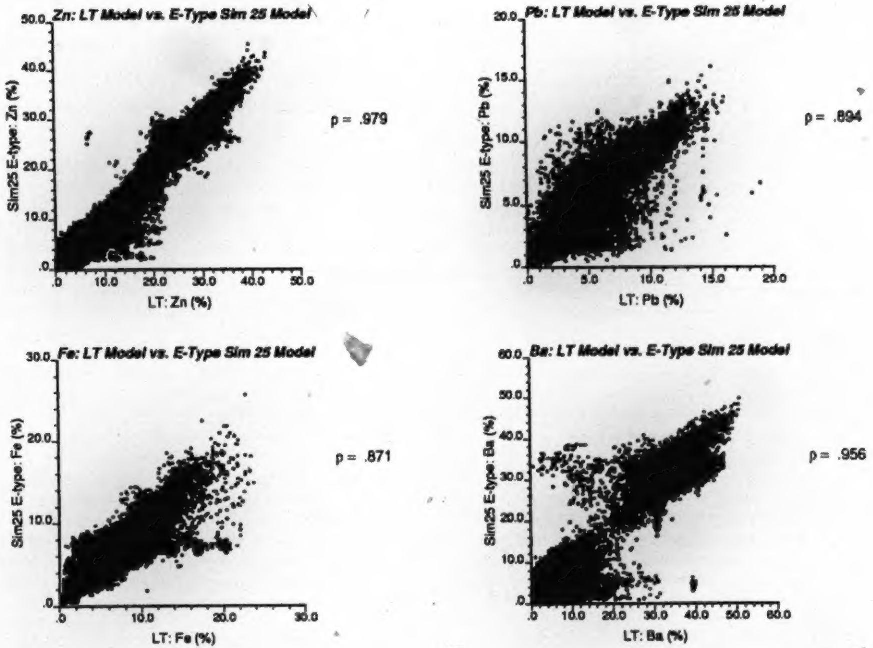


Figure 5.23: Crossplots of Zn, Pb, Fe and Ba from E-type estimates and existing long-term model.

local uncertainty, along with recovery forecasting, resource estimation and uncertainty in short term production.

Note that the applications in this section are for illustrative purposes only; some parameters have been chosen arbitrarily to illustrate the application(s) to be implemented and wherever possible, real functions have been used.

Applications using Local Uncertainty. Multiple realizations allow distributions of uncertainty to be constructed at each location. With these local distributions, different summary statistics can be calculated such as the expected value and probability of interest. Note that the expected value at each location is the E-type estimate that was used in the previous section for model validation. The models that result from these calculations are based on all realizations simultaneously; they are not one realization.

For a cutoff grade of interest, the probability of exceeding this threshold can be assessed using the local distributions from simulation. The use of a high cutoff grade shows areas that are surely high, that is, those areas with a high probability to be high grade. Similarly, a map that shows the probability to be below a low threshold reveals the areas that are almost certainly low. Figure 5.24 shows three probability maps for Zn grade and one for Ba grade. The top two figures shows the reduced area of certainty of finding low and medium grade Zn as a result of increasing the Zn threshold (cutoff grade). The bottom two figures allows for a visual comparison of the region of very high Zn grade ($> 25\%$) and that corresponding to low Ba grade ($< 7\%$). For these maps, the Zn grades were chosen arbitrarily, while the Ba cutoff grade corresponds to the grade specified by the mill for grade control purposes. Since Ba grade adversely affects Zn recovery, it is important to determine the locations within the pit where Ba exceeds the maximum allowable for production. These maps provide one way to quickly determine the general areas where Ba grade may be an issue.

Probability Map of Ore/Waste. For Red Dog, stockpile blending is based on as many as seven different criteria, ranging from grade values of multiple metals, grade ratios between metals, and particle textural criteria. The decision of which material to send to a particular stockpile is initially based on model values, perhaps refined by on-site inspection by mine geologists.

Greater accuracy in the ore/waste classification and stockpile construction can be achieved by using the simulated realizations to determine the transitional zone. Probability maps constructed using the blending criteria would show the transition between ore and waste. Areas of indeterminant probability (0.3 to 0.7) may warrant further sampling.

The methodology to generate such a model is fairly straightforward. The first step is to classify each block within a realization as either ore or waste, and apply a straightforward binary code (e.g. 1=ore, 0=waste). This classification requires taking the first realization for all variables and visiting each block and applying the classification criteria. When all blocks have been visited, the result is an indicator model showing the blocks as either ore or waste. This step is performed for all realizations.

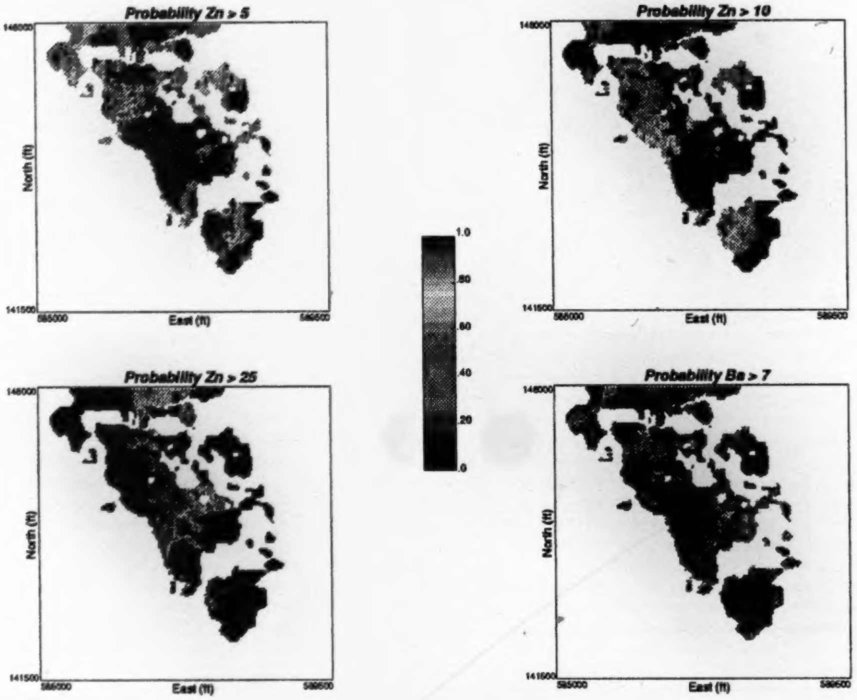

Probability Zn > 5
Probability Zn > 10
Probability Zn > 25
Probability Ba > 7
North (ft)
East (ft)
148000
141500
585000
589500
1.0
.80
.60
.40
.20
.0

Figure 5.24: Probability maps to exceed a specific cutoff: $Zn > 5\%$ (top left), $Zn > 10$ % (top right), $Zn > 25$ % (bottom left), and $Ba > 7$ % (bottom right). The section shown corresponds to bench 850 (spanning elevations 850 to 875ft). Note that the bottom two figures show areas of where the Zn grade is sure to be high (where the probability is close to 1.0) and the corresponding areas where the Ba grade is sure to be low (where the probability is close to 0.0).

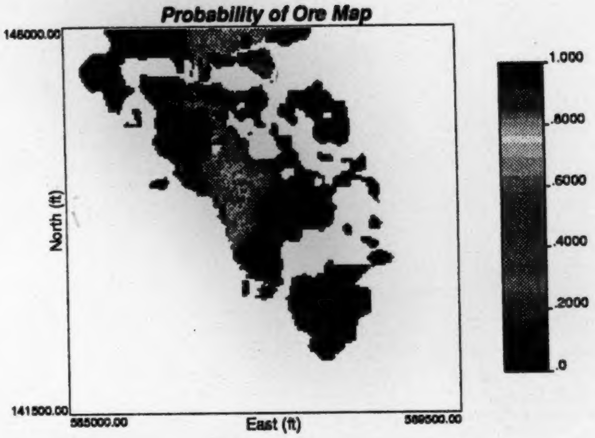


Figure 5.25: Probability of ore map based on stockpile criteria. The section shown corresponds to bench 850 (spanning elevations 850 to 875ft).

The second step involves summarizing the 40 ore/waste models to yield a probability model. This step requires that each block in the ore/waste indicator models is visited (over the 40 realizations), and a simple count is taken of the number of times this block is classified as ore. Divide this number by 40 to yield the probability of ore for this location. This is repeated until all locations have been visited to give a probability of ore model.

The last step is to visualize this probability model (Figure 5.25). The result shows areas that are highly likely to be ore, highly likely to be waste and the transition from one zone to the other. Note that in this case, the stockpile blending criteria, which consists of five different conditions (only grade-based conditions were applied), was used as the classification criteria. These were:

$$\begin{aligned} \text{Zn/Fe ratio} &\geq 2.5 \\ \text{Fe} &\leq 9.0\% \\ \text{Pb} &\leq 5.7\% \\ \text{Ba} &< 7.0\% \\ \text{TOC} &\leq 0.65\% \end{aligned}$$

Satisfaction of the above criteria resulted in a classification of ore. In practice, economic criteria could be used to establish a map of profitability. A block that yields negative profit would be classified as waste, while a block that gives positive profit would be considered ore (see Section 5.6). This would also give a probability of ore map.

Simulating Stockpiles from Models. This application is similar to the previous application. The idea is to apply the blending criteria to specific volumes being planned for a stockpile rather than on each block independently. These volumes will be the construction of one or more stockpiles.

The classification criteria are applied to each of the blocks within the volume over the multiple realizations and multiple variables. A table can be constructed to summarize the grade values from all 40 realizations to assess the mean and variance of the grade distribution for the specific volumes. The probability of ore can be calculated.

Recovery Forecasting. From the onset, a key application of the realizations was to forecast recovery. Of course, this requires an understanding of the metallurgical processes and the effect of metal and contaminant grades on recovery.

The following recovery functions were provided and the conditions were applied in this order of priority:

1. If the material is vein type rock, then a constant Zn recovery of 89 % is applied.
2. For all other rock types:
   (a) If $Ba \geq 7\%$ then Zn recovery is given by

   $$\text{Zn recovery: } 27.182 * ln(Zn) - 3.4834, \text{ to a maximum } 85\%$$

   (b) For $Ba < 7\%$:
      i. If $Fe < 15.5\%$ then

      $$\text{Zn recovery: } 89.4 - 0.7 * Fe$$

      ii. If $Fe \geq 15.5\%$ then

      $$\text{Zn recovery: } (-0.4205*Fe+90.196)-(55-(-0.531*Fe+60.036))*1.6$$

These transfer functions were applied to realizations of multiple grades to calculate the Zn recovery at a specific location. Figure 5.26 shows six realizations of the recovery models generated, while Figure 5.27 shows the maps that correspond to the minimum, average and maximum calculated recovery at each location. The map of minimum local recoveries shows regions that are surely to have high recoveries; the map of maximum local recoveries shows those areas that will surely have low recoveries. Note that in all maps, the areas corresponding to the vein rock unit have a constant recovery factor, in accordance with the above recovery functions.

The result of generating these recovery models is that at each location, a local distribution of uncertainty in the recovery can be constructed (Figure 5.28). Alternatively, consideration of the average recovery based on all locations over the realization would yield the uncertainty distribution in the global recovery (Figure 5.29).

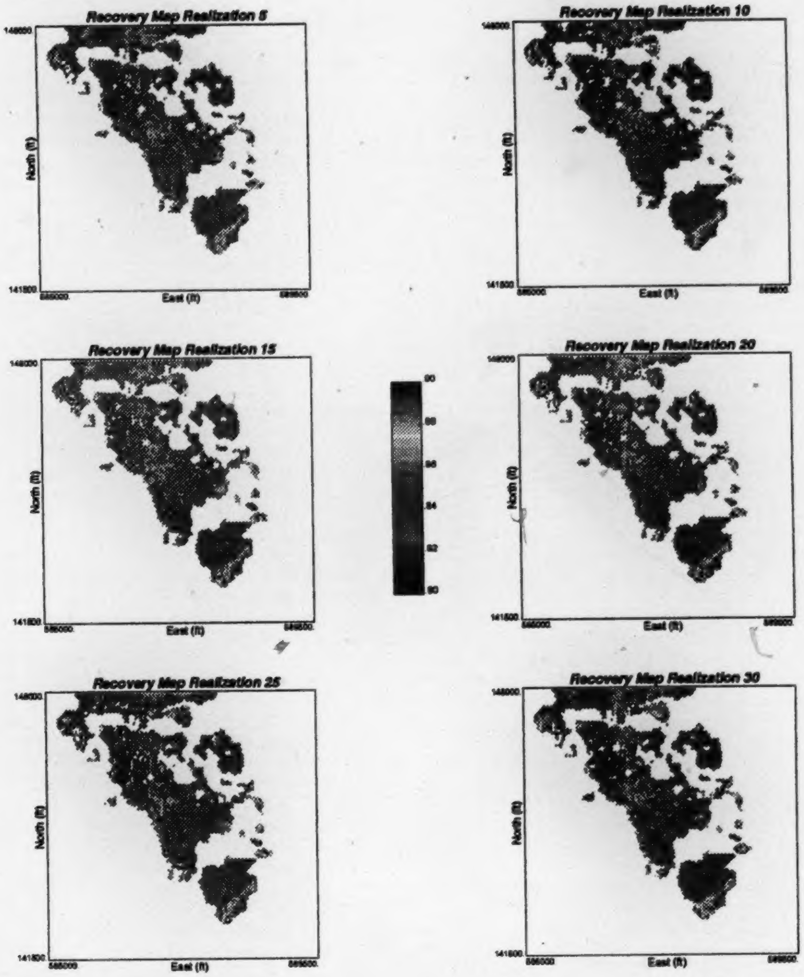

Recovery Map Realization 5
Recovery Map Realization 10
Recovery Map Realization 15
Recovery Map Realization 20
Recovery Map Realization 25
Recovery Map Realization 30
North (ft)
East (ft)
90
88
86
84
82
80

Figure 5.26: Six realizations of the recovery models as calculated based on recovery functions provided by Teck Cominco. The section shown corresponds to bench 850 (spanning elevations 850 to 875ft).

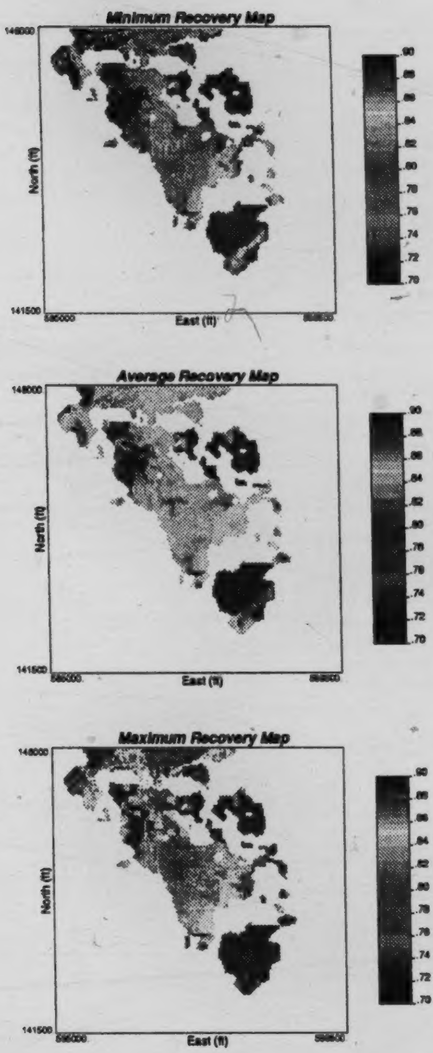

Minimum Recovery Map
148000
141500
North (ft)
East (ft)
.90
.88
.86
.84
.82
.80
.78
.76
.74
.72
.70
Average Recovery Map
148000
141500
North (ft)
East (ft)
.90
.88
.86
.84
.82
.80
.78
.76
.74
.72
.70
Maximum Recovery Map
148000
141500
North (ft)
East (ft)
.90
.88
.86
.84
.82
.80
.78
.76
.74
.72
.70

Figure 5.27: Summary maps of the 40 recovery realizations: the minimum (top), average (middle) and maximum (bottom) recovery at each location. The section shown corresponds to bench 850 (spanning elevations 850 to 875ft).

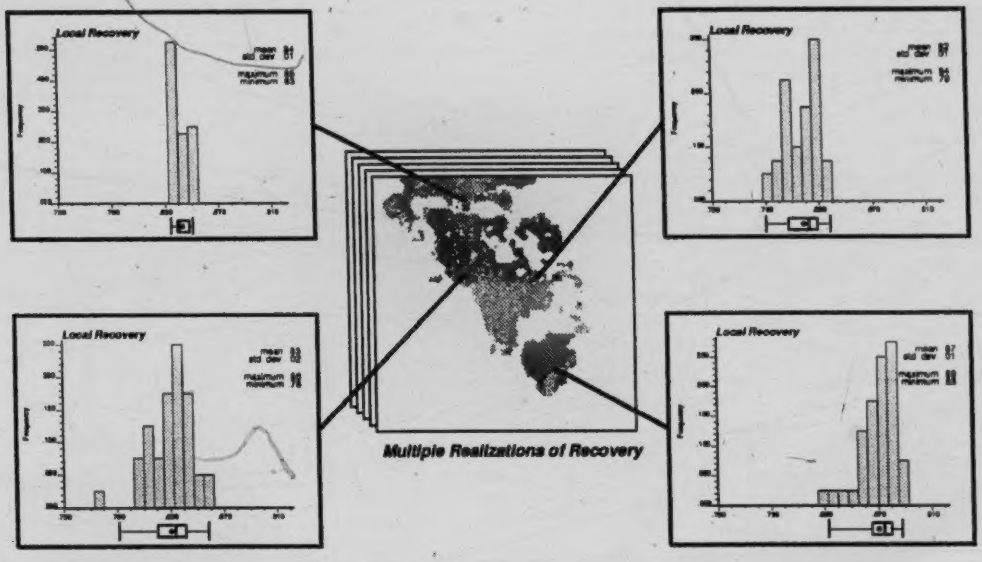


Figure 5.28: Uncertainty in the local recovery is shown for four arbitrarily chosen locations within the model. In all cases, the reference point plotted in the box plot of the histograms corresponds to the mean value.

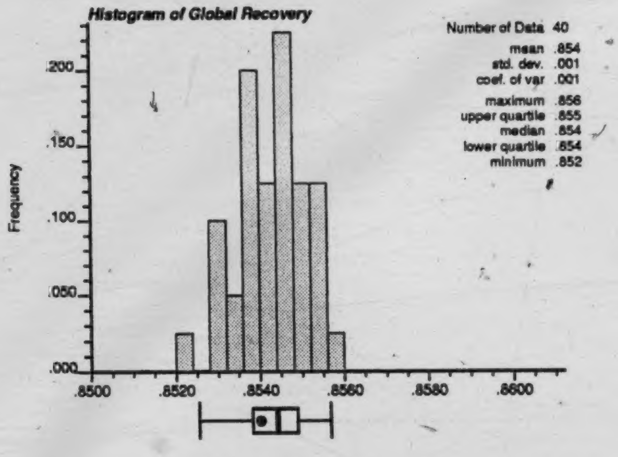


Figure 5.29: Uncertainty in the global recovery based on all 40 realizations of recovery. The reference point plotted in the box plot of the histograms corresponds to the mean value.

Uncertainty in Global Resource. In practice, the global reserve (within an entire pit) is reported as a single number with no indication of the uncertainty in this value. Using multiple realizations, simulation allows for uncertainty assessment of the global reserves.

In the same manner as the recovery models were generated (above), a transfer function to calculate reserves can be applied over a single realization of all variables to determine the reserves based on that realization. This calculation would be repeated for all the 40 realizations to obtain 40 different values for the global reserves. A histogram of these 40 values would show the uncertainty in the reserves.

As the model generated for this case study was only a small portion of the actual mine, and the pit limits were not available, the reserve cannot be determined, however the resource within the model limits can be calculated.

Specific tonnage factor equations were provided by Teck Cominco for the block averaging from the 12.5ft × 12.5ft × 12.5ft to the more practical 25ft × 25ft × 25ft resolution. These equations account for the Zn, Pb, Fe and Ba grades at each block within the grid. As a result, the density for each block within the model limits could be directly calculated.

From the previous application of determining the recovery at each block location, the recoverable resource can be calculated as:

$$\text{recoverable resource} = \text{recovery} * \text{tonnes of material} * \text{Zn grade}/100\%$$

The above equation was applied to each location within the models to determine the available Zn resource. Note again that no economic constraint has been applied (e.g. defined pit limits and/or cutoff grades), so the above calculation is a simple estimate of the material that can be recovered by the mill.

Figure 5.30 shows six realizations of the resource models generated, while Figure 5.31 shows the maps that correspond to the minimum, average and maximum resource estimates at each location. Similar to the assessment of the local recovery, uncertainty in the local resource can be determined at each location (Figure 5.32). Further, uncertainty in the global resource can be assessed by calculating the global resource from multiple realizations and plotting these in a histogram (see Figure 5.33).

Another directly related application is to assess the uncertainty in the resource over a short term period. In this case, the short term period may correspond to monthly or quarterly production, which can be directly traced to a specific volume of material that is planned for mining in the next month or the next quarter. This essentially involves determining the available resource within the specified volume. Figure 5.34 shows an example of this type of application with an arbitrarily chosen volume, and the uncertainty in the available resource is also shown.

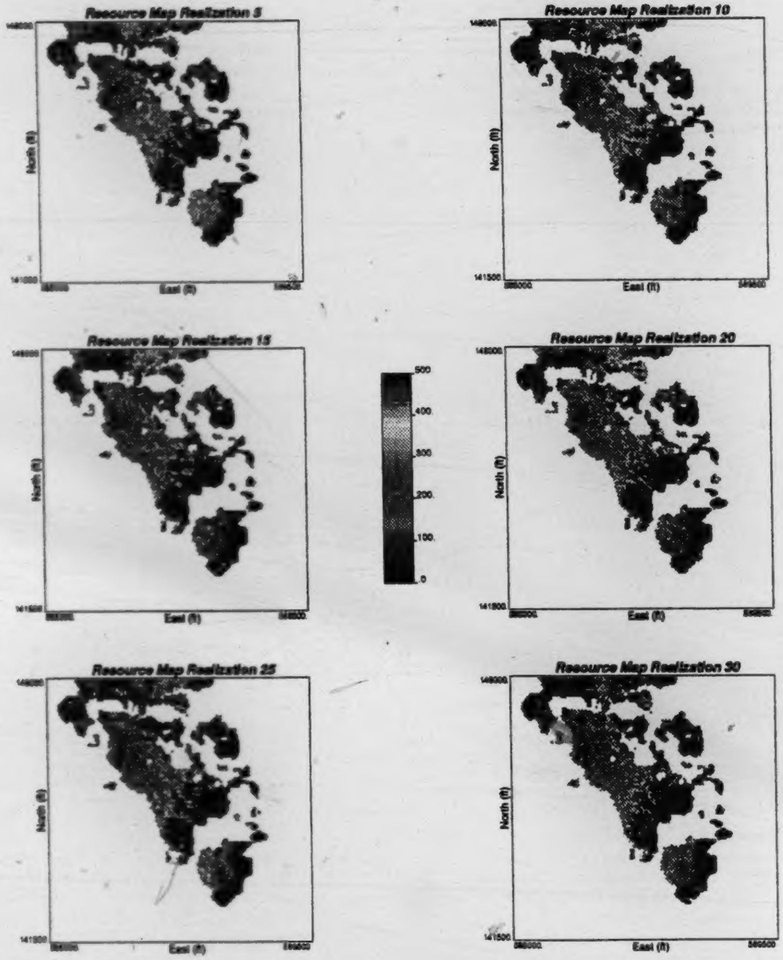

Resource Map Realization 5
Resource Map Realization 10
Resource Map Realization 15
Resource Map Realization 20
Resource Map Realization 25
Resource Map Realization 30
North (ft)
East (ft)
146000
141500
500
400
300
200
100
0

Figure 5.30: Six realizations of the resource models as calculated based on tonnage factors and recovery functions provided by Teck Cominco. The section shown corresponds to bench 850 (spanning elevations 850 to 875ft).

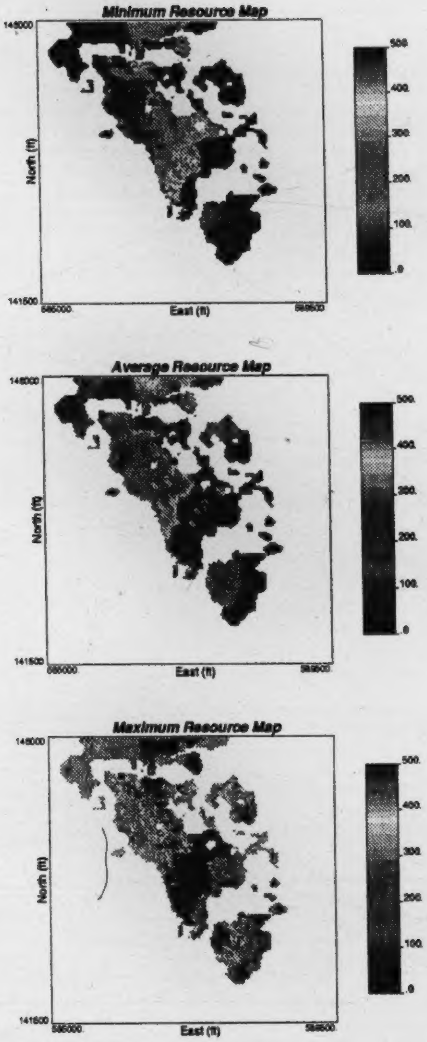

Minimum Resource Map
145000
141500
North (ft)
East (ft)
585000
588500
500.
400.
300.
200.
100.
0
Average Resource Map
145000
141500
North (ft)
East (ft)
585000
588500
500.
400.
300.
200.
100.
0
Maximum Resource Map
145000
141500
North (ft)
East (ft)
585000
588500
500.
400.
300.
200.
100.
0

Figure 5.31: Summary maps of the 40 resource realizations: the minimum (top), average (middle) and maximum (bottom) resource map at each location. The section shown corresponds to bench 850 (spanning elevations 850 to 875ft).

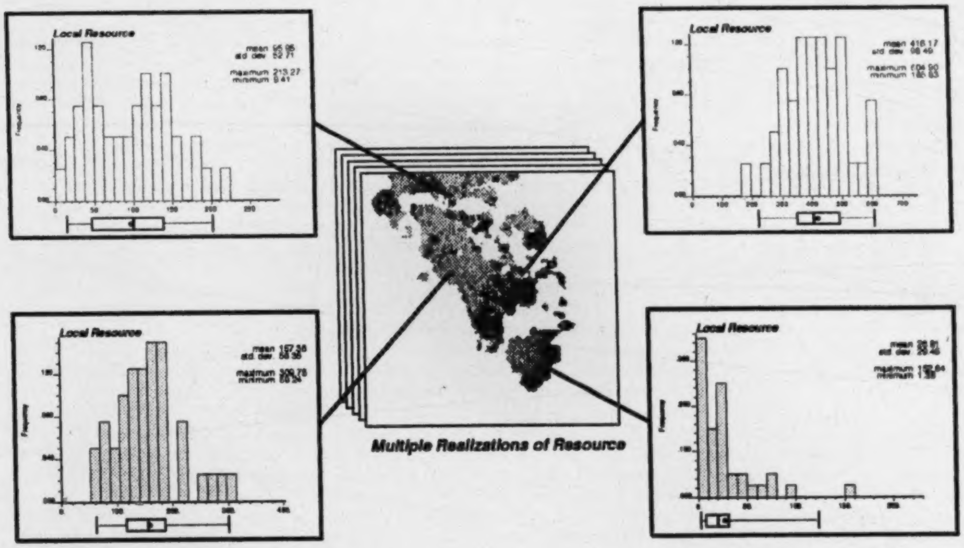


Figure 5.32: Uncertainty in the local resource is shown for four arbitrarily chosen locations within the model (same locations as shown in Figure 5.28). In all cases, the reference point plotted in the box plot of the histograms corresponds to the mean value.

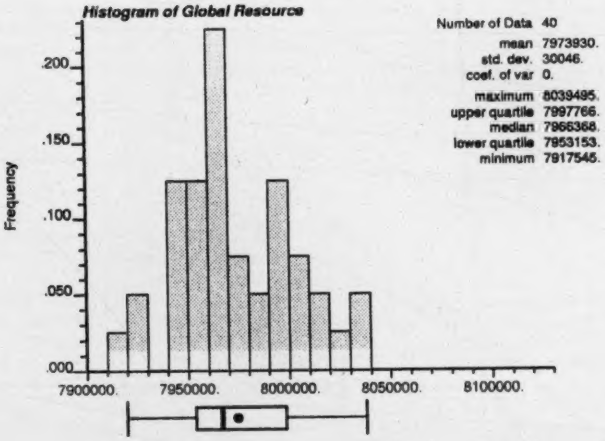


Figure 5.33: Uncertainty in the global resource based on 40 realizations. The reference point plotted in the box plot of the histogram corresponds to the mean value.

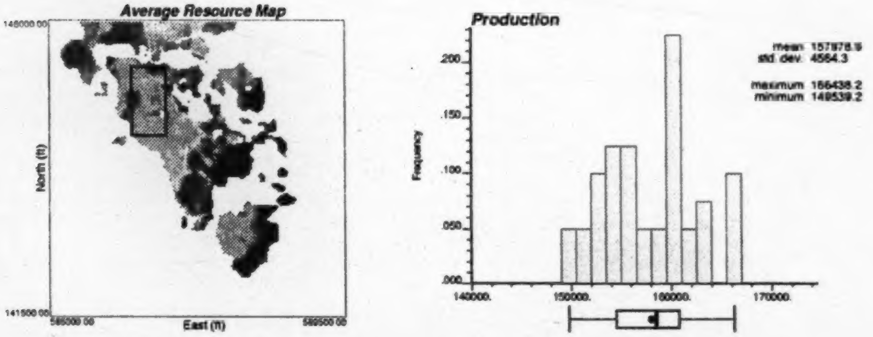


Figure 5.34: Illustration of application for short term planning. The volume of material associated to the planned production for one month is shown on the left, and the uncertainty in the resource available is shown on the right. The reference point plotted in the box plot of the histogram corresponds to the mean value.

5.6 The Value of Simulation

As the previous section showed, there are many possible applications of simulation. In practice, multiple variables are estimated independently with ordinary kriging. This section addresses the impact of the multivariate simulation approach using the stepwise conditional transform relative to the conventional practice of kriging.

The idea is to compare the profit of ore from both methods with true reference data coming from Red Dog. A profit function is applied to obtain a true profit dataset. A subset of the reference data will be extracted and used to model the grades using both kriging and simulation. The profit function will be applied to these grade models. Based on the expected profit from each approach, each block within the model will be classified as either ore or waste. The true profit at each location is known, so the profit from each model can be calculated.

Profit Function. The real profit function was not available; a profit function was developed for this exercise. The following simple function was proposed:

$$profit(\mathbf{u}) = (Zn(\mathbf{u}) \cdot r_z \cdot f_1(Ba(\mathbf{u})) \cdot f_2(Fe(\mathbf{u})) \cdot p_z + Pb(\mathbf{u}) \cdot r_p \cdot p_p - c_{fix}) \cdot tons \quad (5.2)$$

where

$\mathbf{u}$ = location vector
Zn = Zn grade
r_z = Zn recovery function used to scale the maximum Zn recovery
$f_1(Ba)$ = factor that accounts for effect of Ba on Zn recovery
$f_2(Fe)$ = factor that accounts for effect of Fe on Zn recovery
p_z = price of Zn, in $US/ton
Pb = Pb grade
r_p = Pb recovery
p_p = price of Pb, in $US/ton
c_{fix} = fixed cost in $US/ton
$tons$ = tons of material based on specific gravity equations provided by Teck Cominco

The only information available are the metal grades. All other parameters were developed or chosen to be constant. The metal recoveries for both Zn and Pb, r_z and r_p, were calculated as Red Dog's five year average recovery (1998-2002) based on Teck Cominco's financial report [78]. These were 83.6% Zn recovery and 58.7% Pb recovery. The price for Zn was chosen to be $680/ton of Zn, and the price for Pb was chosen as $380/ton of Pb; both prices were approximated based on the metal prices from the London Metal Exchange in 2003 [79].

Although recovery functions were provided by Teck Cominco (Section 5.5), those functions did not actually depend on Ba grade. At the time of this work, Teck Cominco was developing new functions based on extensive metallurgical testing. In light of this lack of confidence in the recovery functions, the Zn recovery function was scaled by functions that quantify the impact of Ba grade and Fe grade as a

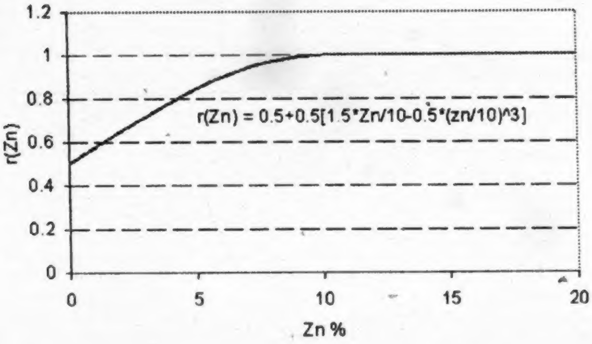


Figure 5.35: Zn recovery function developed for comparison of kriging and simulation. Function for grades between 0 and 10% is shown, no reduction in recovery was expected beyond 10%.

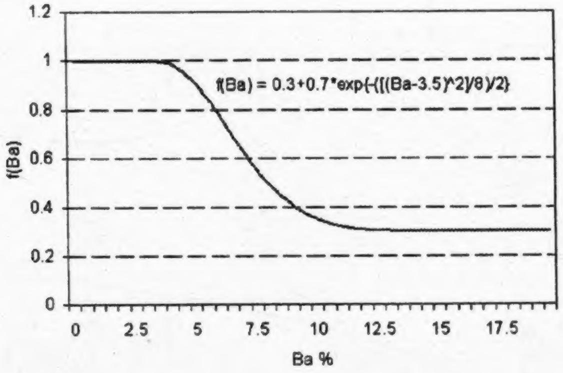


Figure 5.36: Discount function on Zn recovery due to %Ba content.

fraction of the maximum recovery. For Zn, the recovery function (see Figure 5.35) reaches a constant maximum recovery beyond a threshold grade of 10% Zn. Lower Zn grades than this threshold results in a fraction of the maximum recovery, to a minimum of 50%. The rate of this change was expected to be fairly gradual.

A discount function for Ba content was developed by considering that a threshold grade of 7% Ba resulted in significant impact on Zn recovery. At grades below this threshold, the discount factor was expected be fairly constant. Near the threshold grade of 7%, an inflection point in the discount function was expected, and would gradually flatten at a minimum discount of 35% since some Zn would still be recovered (Figure 5.36).

The discount function for Fe was based on the recovery functions provided by Teck Cominco (Section 5.5), and is illustrated in Figure 5.37.

All that remains to determine is the fixed cost. For this, an arbitrary cost per ton mined was chosen such that the area of interest yielded approximately 50% ore

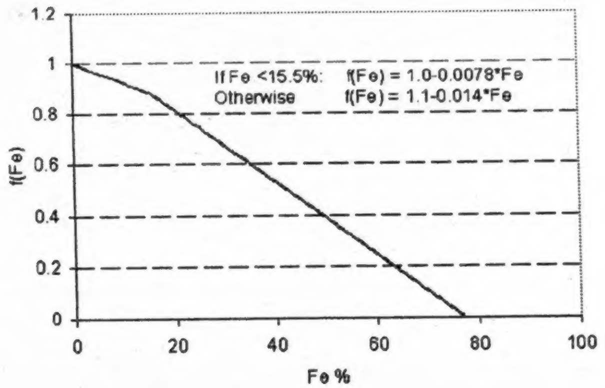


Figure 5.37: Discount function on Zn recovery due to %Fe content. Two different discount functions were applied with a threshold grade of 15.5% Fe. These functions were based on the recovery functions provided by Teck Cominco Limited.

and 50% waste classification. This depends on the region chosen for modeling; for this exercise and the area described below, the fixed cost was set at $128/ton and accounts for all operational costs including mining and milling cost.

Reference Data. For a fair comparison to be made, real data must be used. The density and number of BH data available make it an attractive database for true data. Rather than modeling the entire area, only a small area will be modeled. The area was chosen to be in a marginal zone, where ore/waste classification based on the models would have the largest impact.

Figure 5.38 shows the available BH data in the chosen region of 400ft × 400ft in the 850 bench (spanning elevations 850 to 875ft), and the subset of data extracted from this region. The available data consists of 532 BH samples of Zn, Pb, Fe and Ba. From this dataset, 25 samples separated at a nominal 100ft × 100ft spacing were chosen to act as exploration data. This spacing is consistent with the DH data available for Red Dog. This subset of data was used as conditioning data for kriging and simulation.

Model Construction. The model grid was chosen to be 10ft × 10ft × 25ft, which is similar to the 10ft × 12ft × 25ft spacing of the BH data. A total of 1600 blocks were modeled.

With only 25 samples available for modeling, variography would be very difficult. To filter out the influence of poor variogram inference, variograms for both approaches were calculated and fitted using the reference 532 BH data.

The variograms for kriging were calculated for the original data (Figure 5.39). The variograms for simulation were calculated and fitted for the stepwise conditionally transformed data (Figure 5.40). In both sets of variograms, a trend was apparent from the experimental points extending beyond the sill of 1.0. This was not surprising given that the area was purposely chosen to be in the transition zone

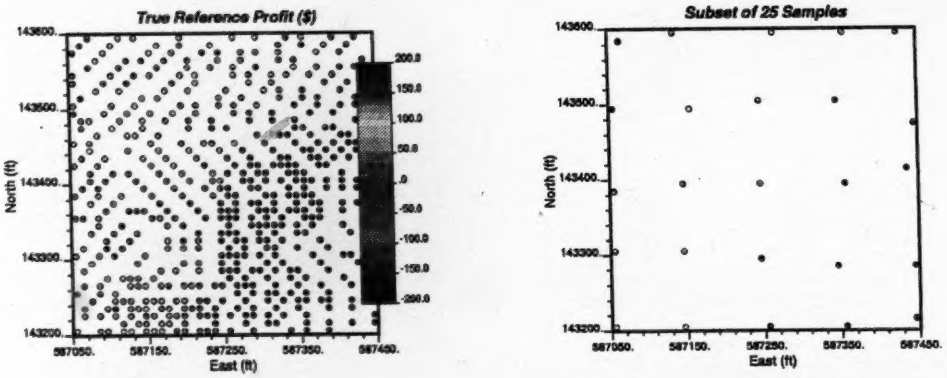

True Reference Profit ($)
143600.
143500.
143400.
143300.
143200.
North (ft)
587050.
587150.
587250.
587350.
587450.
East (ft)
200.0
150.0
100.0
50.0
0
-50.0
-100.0
-150.0
-200.0
Subset of 25 Samples
143600.
143500.
143400.
143300.
143200.
North (ft)
587050.
587150.
587250.
587350.
587450.
East (ft)

Figure 5.38: Location map of reference BH data (left) and sampled BH data (right) for use in comparing model approaches.

between ore and waste material, hence a trend from low to high grades was expected. Trend modeling was not performed for this exercise because of the relatively small area.

For kriging, each variable was estimated independently using ordinary kriging. For simulation, the stepwise conditionally transformed variables were independently simulated using sequential Gaussian simulation to generate 100 realizations of the grades. Figure 5.41 shows a comparison of the estimated grades from kriging and one realization of the simulated grades. As expected, the kriged models were very smooth, while the simulated realization showed greater variability while honouring the same large scale features shown in the kriged models.

Results. These grade models were then processed by applying the profit function at each location within the model. Although, 100 realizations of profit were available from simulation, the ore/waste classification was based on the expected profit map obtained by calculating the expected value of profit at each location. Figure 5.42 shows the profit map obtained from simulation and kriging along with the true profit at the 532 locations where real data were available.

Although 1600 locations were modeled, only the 532 blocks corresponding to locations where true data were available can be compared. At these locations, the true profit was known. The models from kriging and simulation were used to classify the 532 locations as either ore or waste:

$$i(\mathbf{u}_\alpha; profit) = \begin{cases} ore, & \text{if profit}(\mathbf{u}_\alpha) \geq 0 \\ waste, & \text{if profit}(\mathbf{u}_\alpha) < 0 \end{cases}$$

Figure 5.43 shows the comparison of the ore/waste classification of the 532 locations from the true relative to the kriging and the simulation approaches. Overall, both approaches clearly show the waste and the ore region; relatively few blocks were misclassified.

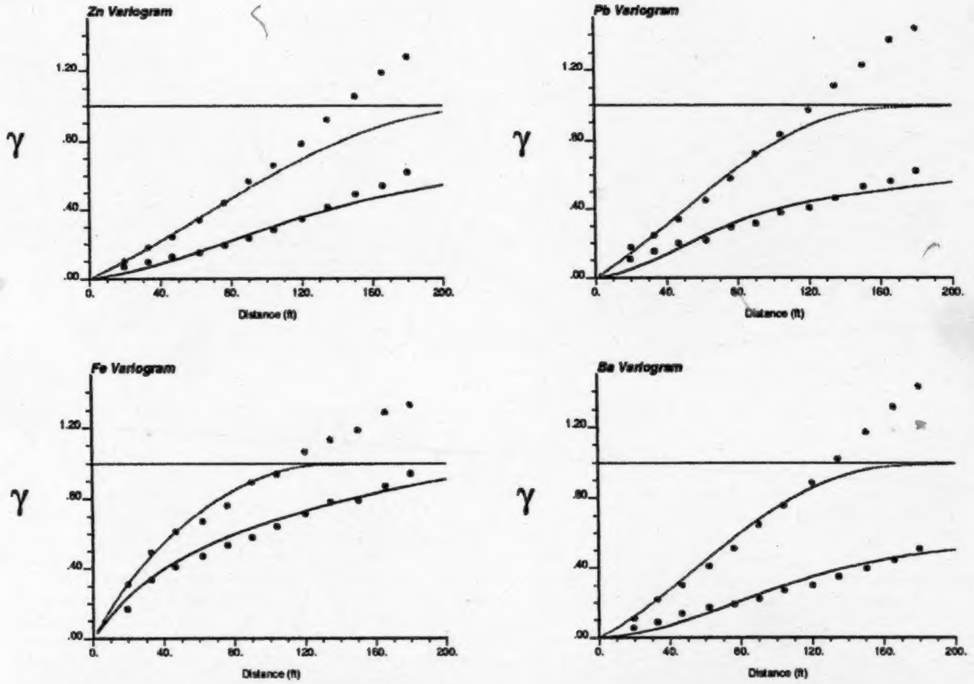


Figure 5.39: Variograms of direct space data for use in ordinary kriging approach: Zn (top left), Pb (top right), Fe (bottom left) and Ba (bottom right). The two directions shown correspond to the horizontal minimum and maximum directions of continuity.

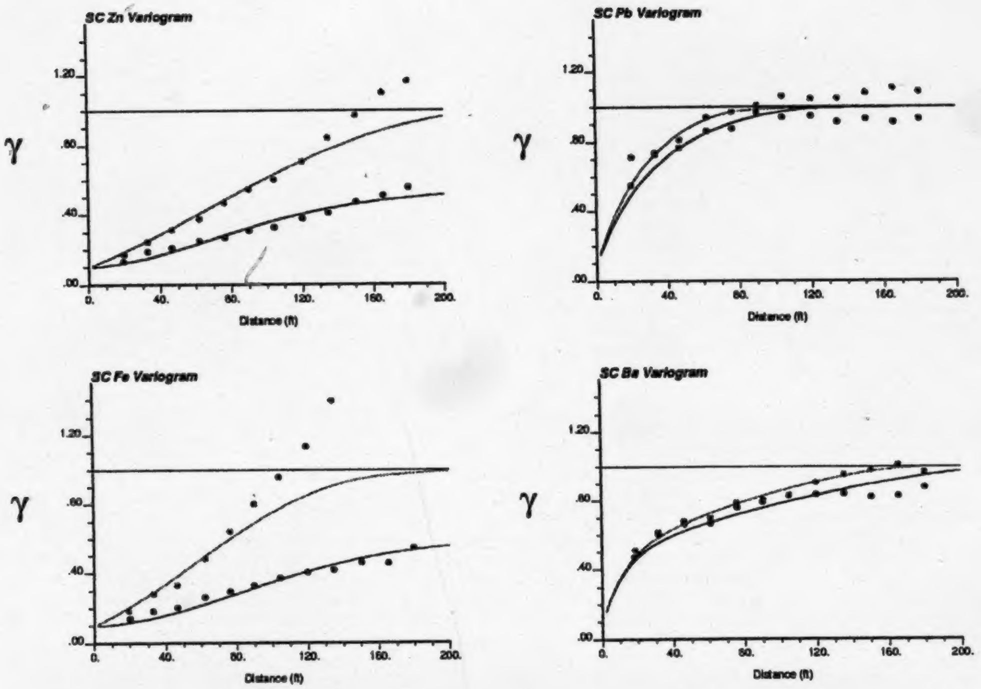


Figure 5.40: Variograms of stepwise conditional scores for use in simulation approach.: Zn (top left), Pb (top right), Fe (bottom left) and Ba (bottom right). The two directions shown correspond to the horizontal minimum and maximum directions of continuity.

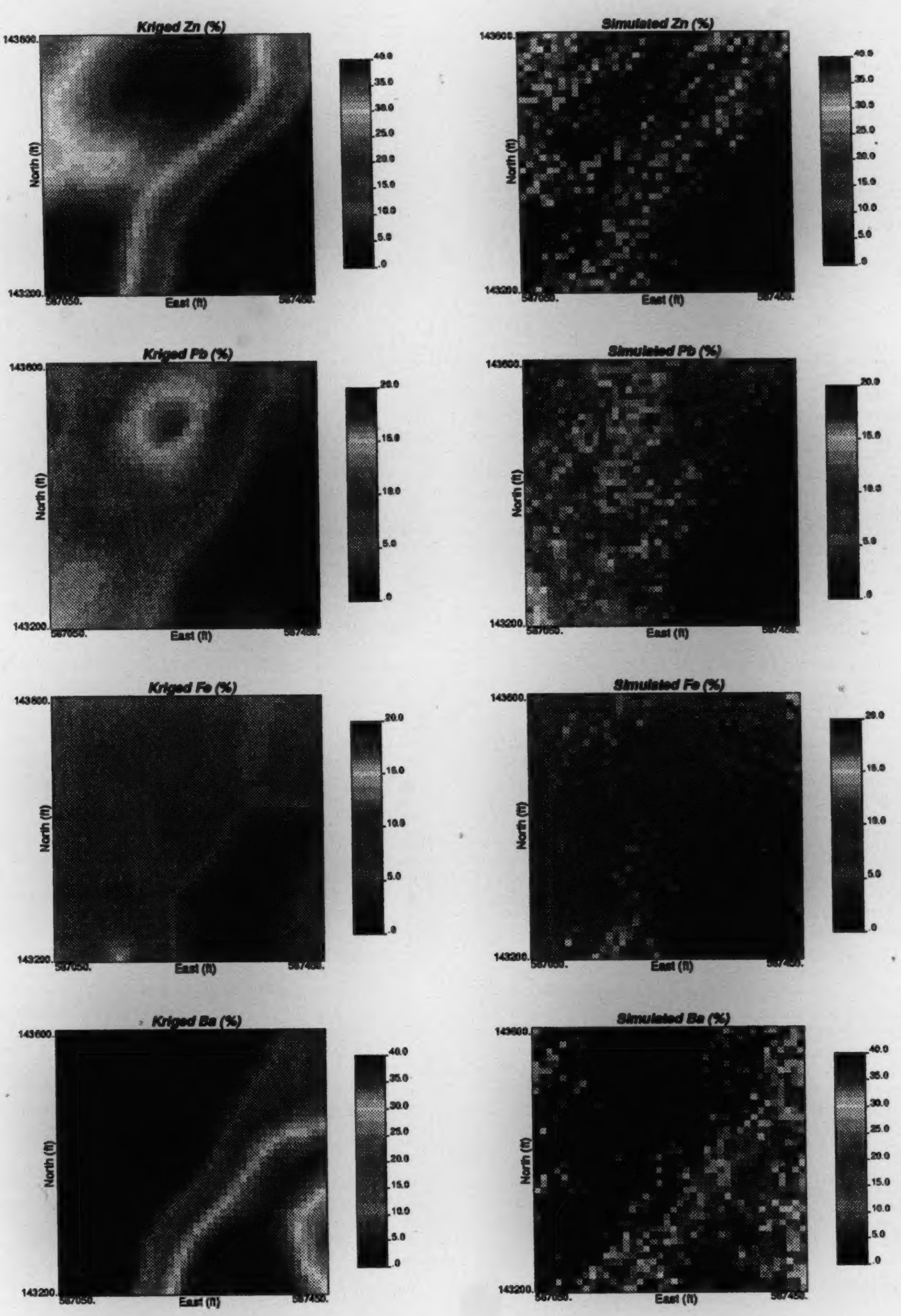
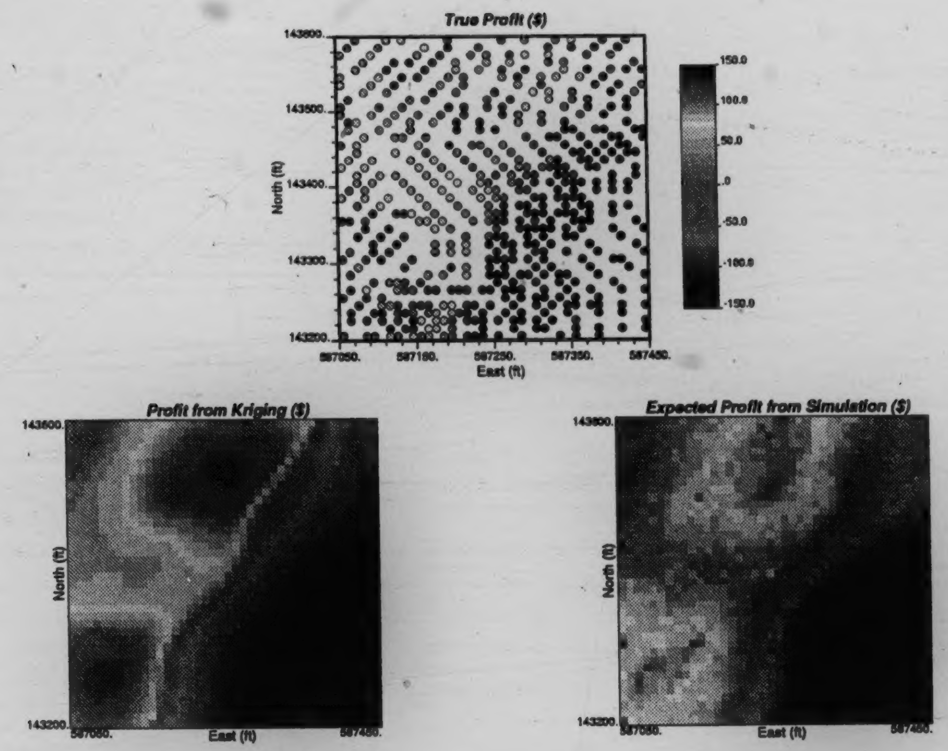

Kriged Zn (%)
Simulated Zn (%)
Kriged Pb (%)
Simulated Pb (%)
Kriged Fe (%)
Simulated Fe (%)
Kriged Ba (%)
Simulated Ba (%)
North (ft)
East (ft)
143600.
143200.
587050.
587450.
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0

Figure 5.41: Comparison of kriged model (left) and one realization from simulation (right) for Zn, Pb, Fe and Ba (from top to bottom).


True Profit ($)

Profit from Kriging ($)

Expected Profit from Simulation ($)

143600.
143500.
143400.
143300.
143200.
North (ft)
587050.
587150.
587250.
587350.
587450.
East (ft)
150.0
100.0
50.0
.0
-50.0
-100.0
-150.0
143600.
143200.
587050.
East (ft)
587450.
North (ft)
143600.
143200.
587050.
East (ft)
587450.
North (ft)

Figure 5.42: Comparison of true profit map at data locations (top) and the profit map for ore/waste classification from kriging (bottom left) and simulation (bottom right).

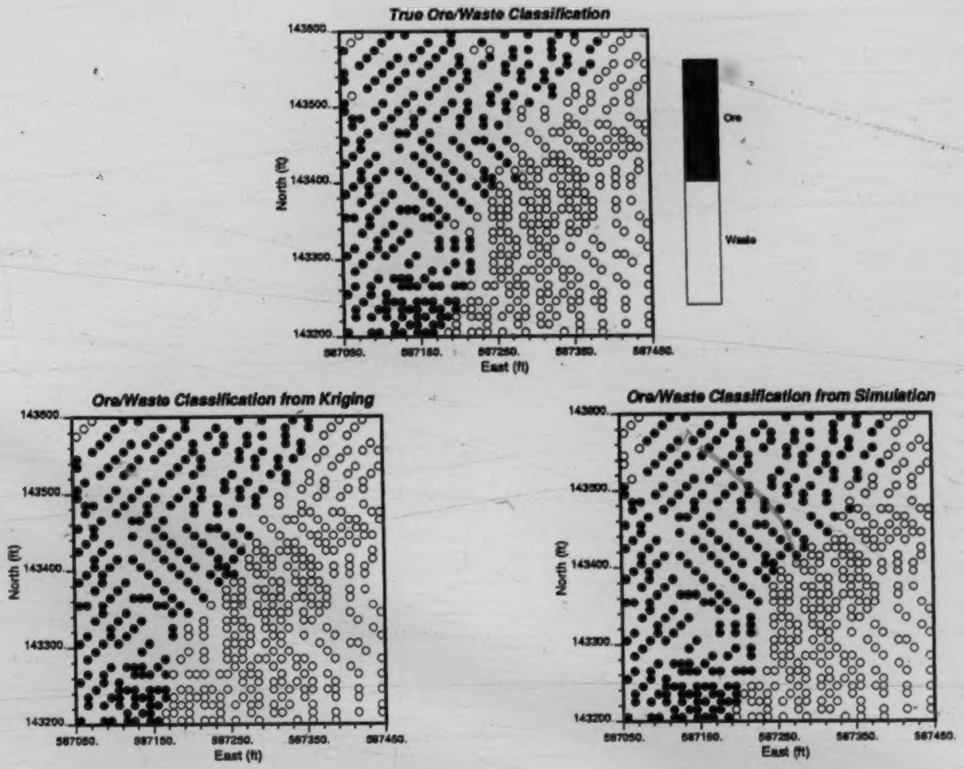


Figure 5.43: Comparison of true ore/waste classification (top) and the classification from kriging (bottom left) and simulation (bottom right) at data locations.

Kriging	TRUE	
	Ore	Waste
Ore	225	11
Waste	26	270

Simulation	TRUE	
	Ore	Waste
Ore	246	27
Waste	5	254

Figure 5.44: Ore/Waste classification summary of kriging (left) and simulation (right) relative to true ore/waste classification.

Figure 5.44 shows the summary of the ore/waste classification from both kriging and simulation relative to the true classification. The tables show that the kriging approach resulted in a total 7% of blocks that were misclassified, compared to the 6% misclassified by simulation. From the true profit, 251 blocks (47% of the true data) were classified as ore; simulation correctly classified ore for 98% of those blocks while kriging correctly classified 90% of those blocks.

For those blocks classified as ore, the profit of ore mined as a result of the classification from each method was compared with the true profit of $7.89M (million). The results from such a comparison showed that the simulation approach yielded $7.28M while kriging yielded $7.06M in profit. Although these profit values appear high for the relatively small area, the relative percentage increase in profit is the key result. Multivariate simulation resulted in 92% of the true profit relative to the 89% yielded by kriging. In practice, this 3% difference may translate to several millions of dollars in increased profit.

5.7 Remarks

For the seven variables within the eight rock types, conditional simulation models were constructed using the stepwise conditional transformation technique to account for the multivariate relations. 12.5ft composites and a geology model at $(25\text{ft})^3$ resolution were used to develop these models. Each model was validated by checking reproduction of the input drillhole data, representative histogram, variogram, and the multivariate distributions.

Validation with additional data and comparisons to the existing long term model showed the conditional simulations have similar predictive abilities to the existing models. Multivariate simulation provides two significant improvements from the existing long term model. Firstly, the simulated realizations account for the complex multivariate relations inherent in the data, resulting in models that respect these relations on both a local and global scale. Secondly, the simulation models provide a basis for some interesting applications for decision making and risk assessment. These applications range from classification of ore/waste regions based on complex criteria to recovery forecasting given a clear understanding of metallurgical processes and relations.

A comparison of the multivariate simulation approach used in this case study and the common practice of kriging multiple variables independently showed that the simulation models resulted in an increase in profit of 3% over the kriging approach, yielding a total of 92% of the true profit.

Chapter 6

Comparison of Multivariate Cosimulation Algorithms

There are only a few cosimulation algorithms that are applied in practice. A balance is struck between honouring the available information and the simplicity of the modeling process. The most complex step in geostatistical modeling is the choice and subsequent fitting of a coregionalization model. The type of simulation that follows is a direct consequence of this decision.

Full cosimulation results from the adoption of a linear model of coregionalization (LMC), while the simpler (and more common) collocated cosimulation is a consequence of adopting the Markov hypothesis. Commercial and public domain software implement the conventional simulation algorithms, so the actual simulation step of a project is relatively straightforward.

This chapter compares the results of conventional cosimulation algorithms to a real petroleum dataset. The techniques employed are a direct consequence of adopting some of the models of coregionalization (Section 2.1.4). Four practical methods are compared: (1) cosimulation with full cokriging, (2) cosimulation with collocated cokriging, (3) cosimulation with stepwise conditionally transformed variables, and (4) indicator simulation with the Markov-Bayes model. Multiple realizations are generated for each technique and then processed through a simple flow simulation. Each method is compared based on the resulting flow performance, as well as reproduction of the multivariate distribution.

Unfortunately, we have no reference true values so all we can do is note the significant differences in the results and encourage careful application of multivariate geostatistical tools.

6.1 The Data

The dataset is made available by Amoco. There are three variables of interest: porosity, permeability and seismic data. The seismic data is considered as soft data that is available in 2-D (see Figure 6.1), which will be used as secondary data in the cosimulation of porosity and permeability. Data sampling is fairly regular (Figure 6.1), with large areal extents and relatively small thickness. A 2-D simulation

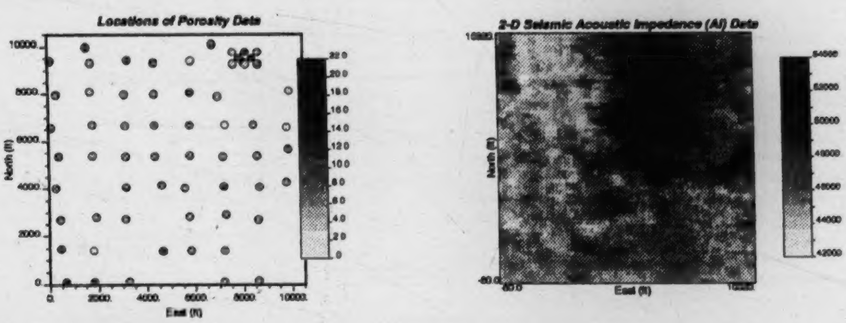

Locations of Porosity Data
10000.
8000.
6000.
4000.
2000.
0.
North (ft)
0.
2000.
4000.
6000.
8000.
10000.
East (ft)
22.0
20.0
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0
2-D Seismic Acoustic Impedance (AI) Data
North (ft)
East (ft)

Figure 6.1: Location map of available hard data (left). Only porosity is shown (collocated permeability data is available). Map of available seismic data for use as soft data (right).

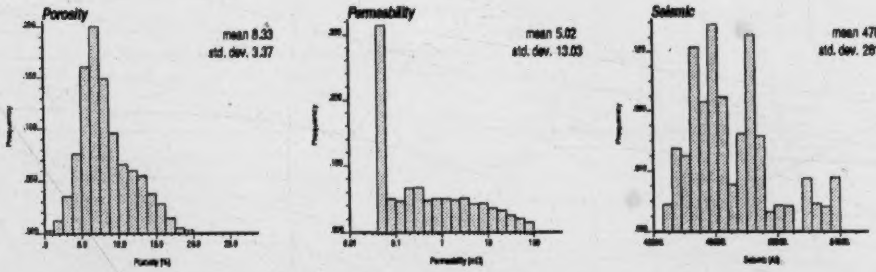


Figure 6.2: Histogram of porosity (left), permeability (centre) and collocated seismic data (right).

exercise will be considered.

Porosity and permeability are hard data and represent the two variables that will be simulated. The distributions for each variable are shown in Figure 6.2, and the bivariate relations between all three variables are presented in Figure 6.3. It is clear that a functional relationship with porosity was used to obtain permeability data.

In the subsequent sections, multiple cosimulation algorithms will be applied to this dataset. The coregionalization models used in each technique are different, so the variogram model(s) used in simulation will be presented in the appropriate simulation section.

6.2 Cosimulation Algorithms

All of the techniques share one feature: all employ a sequential simulation approach. Recall from Section 2.1.7 that the basic steps in sequential simulation, excluding any prior transforms, are:

1. Define a random path visiting each location in the domain.

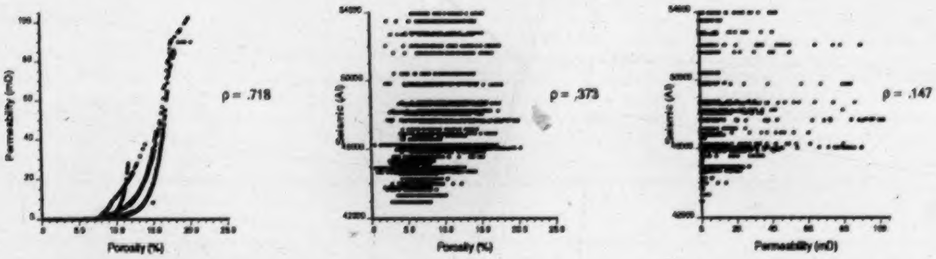


Figure 6.3: Cross plot of porosity-permeability (left), porosity-seismic (centre) and permeability-seismic data (right).

2. At each location:
    (a) Search for nearby primary (and secondary) data, and previously simulated nodes.
    (b) Perform (co)kriging to determine the parameters of the conditional cumulative distribution function (ccdf).
    (c) Draw from the ccdf, defined by the kriged parameters, using Monte Carlo simulation.
    (d) Proceed to next location, and repeat until all locations are visited.

Both the conventional full and collocated cosimulation follow the above procedure almost exactly (after normal score transformation) - the main differences lie in the number of secondary data used in the cokriging step and the effort required to infer the full LMC. These two approaches follow directly from adopting either an LMC or a Markov Model (Section 2.1.4), respectively.

The stepwise conditional transformation to improve multivariate Gaussian simulation presents yet another cosimulation approach. A complex model of coregionalization is an implicit result of applying this multivariate transform (Section 3.1); however, the resulting independent Gaussian variables only require independent simulation, making the approach simple to apply.

Alternatively, a sequential indicator approach may be used to estimate the ccdf directly, rather than assuming a simple parametric distribution (Section 2.1.7). The Markov-Bayes coregionalization model permits the consideration of soft data, such as seismic impedance.

These four cosimulation algorithms will be compared. Details of the variogram modeling are provided in the next sections. For all techniques, only a 2-D horizontal simulation at mid-depth was performed. One hundred realizations were generated for each variable. In each simulation, the entire dataset was used for 3-D variogram modeling, but only the samples at the midpoint in each well were used to condition the simulation. These conditioning data were *not* assigned to grid nodes.

Cosimulation using Full Cokriging. Since simulation will be performed in normal or Gaussian space, a normal score transform was independently applied to each

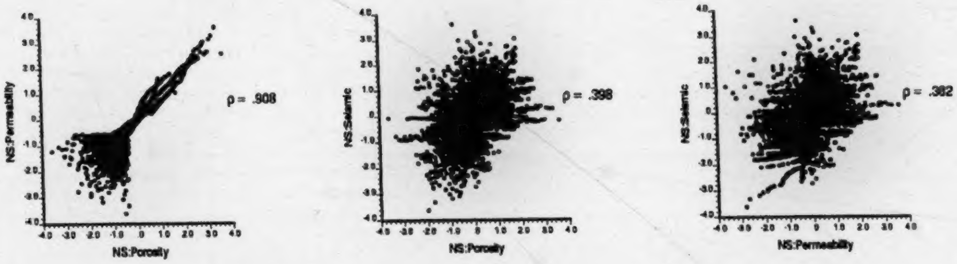


Figure 6.4: Cross plot of normal scores for porosity-permeability (left), porosity-seismic (centre) and permeability-seismic data (right).

Variable-Variable	Nugget	cc Structure 1	cc Structure 2	cc Structure 3
Type		Exponential	Spherical	Spherical
Range Parameters		$a_v = 20$ $a_{hmax} = 3000$ $a_{hmin} = 2000$	$a_v = 70$ $a_{hmax} = 20000$ $a_{hmin} = 5000$	$a_v = 70$ $a_{hmax} = 38000$ $a_{hmin} = 10000$
$\phi - \phi$	0.05	0.40	0.30	0.25
$K - K$	0.05	0.50	0.24	0.21
$AI - AI$	0.00	0.40	0.45	0.15
$\phi - K$	0.05	0.44	0.20	0.22
$\phi - AI$	0.00	0.10	0.11	0.19
$K - AI$	0.00	0.00	0.26	0.12

Table 6.1: Table of variance contribution (cc) of each nested structure in the six variograms constituting an LMC model for full cokriging.

variable. The resulting crossplots are shown in Figure 6.4, and clearly show non-Gaussian features.

Challenges in modeling an LMC for more than two variables make this approach cumbersome to employ. Nevertheless, an LMC model is required to define the spatial relationships between the hard porosity and permeability data to the soft seismic data. The six experimental and model variograms for the horizontal direction are shown in Figure 6.5, and are tabulated in Table 6.1. As the seismic data is 2D, no vertical variograms were calculated or modeled for this variable. The direct and cross variograms in the vertical direction for porosity and permeability are shown in Figure 6.6.

Cosimulation using Collocated Cokriging. This alternative is attractive because fewer variograms have to be calculated and fitted. Note that only Markov Model I will be applied (henceforth simply referred to as the Markov Model). This model only requires that the normal score variogram for the primary data be modeled, and the collocated secondary data be used in cokriging.

Again Gaussian simulation will be performed, so a normal score transform was applied to porosity and permeability. The correlation coefficients in the cross plots shown in Figure 6.4 were required for this simulation. The required normal scores variograms are the same as the direct variograms listed in Table 6.1.

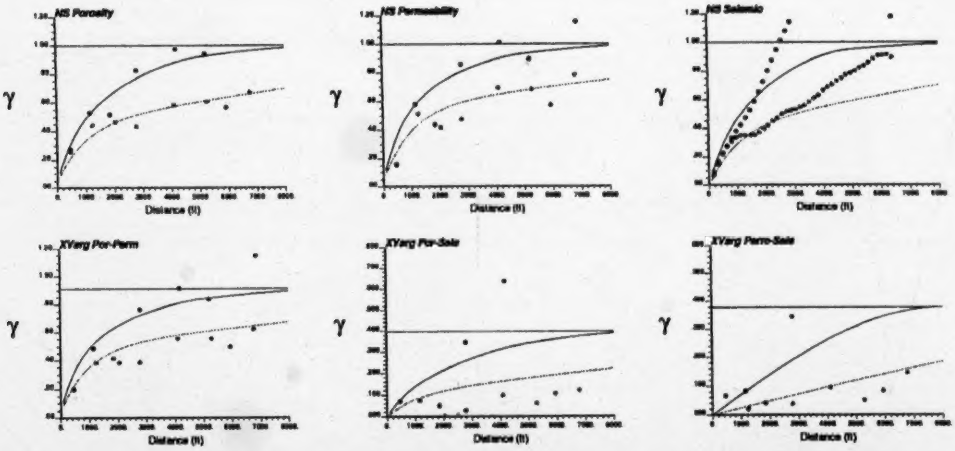


Figure 6.5: Direct variograms for porosity (top left), permeability (top centre) and seismic data (top right). Cross variograms between porosity-permeability (bottom left), porosity-seismic (bottom centre) and permeability-seismic (bottom right). Solid lines are the fitted models for the experimental points shown. Dark lines correspond to the direction of minimum horizontal continuity, while lighter lines correspond to direction of maximum horizontal continuity.

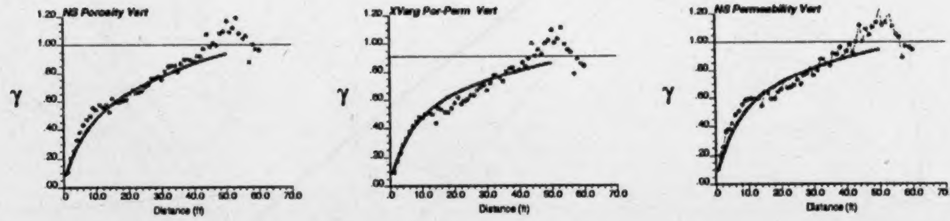


Figure 6.6: Vertical variograms for Amoco data: direct variogram for porosity (left), cross variogram between porosity and permeability (centre) and direct variogram for permeability (right). Solid lines are the fitted models for the experimental points shown.

Stepwise Conditional Transform. From Figure 6.4, the trivariate distribution of the normal scores was clearly non-Gaussian. Applying Gaussian simulation using the normal scores would ignore this non multi-Gaussian behaviour. For cases such as these, the stepwise conditional transform can be applied.

Since seismic data was available at all locations in the 2-D grid, it was selected as the primary variable. Porosity was chosen as the secondary variable and permeability was the last variable to be transformed (conditional to both seismic and porosity data). With over 3000 data available, ten probability classes were chosen, resulting in a minimum of 30 data to define each conditional distribution.

The multivariate distribution of the stepwise conditional (SC) scores are shown in Figure 6.7. A banding effect is apparent. This is partially attributed to the fact that seismic data is 2-D, and so the seismic data that is collocated with the data is a constant at each well. Secondly, the strong functional relationship between porosity and permeability (in Figure 6.3) may have been transferred to the transformed Gaussian space.

Only two variograms were required to be modeled - SC porosity and SC permeability (since seismic was available everywhere, there was no need to simulate this variable). The cross variograms for $\mathbf{h} > 0$ between the transformed variables were checked, with a result showing a maximum correlation of 0.07. Since this correlation was quite low, independence of these variables was assumed. The direct variograms for SC porosity and SC permeability are shown in Figure 6.8 fitted with the following model:

$$\gamma_{\phi}(\mathbf{h}) = 0.65Exp \underset{\substack{a_{hv}=14\\ a_{hmax}=1800\\ a_{hmin}=1500}}{}(\mathbf{h}) + 0.35Sph \underset{\substack{a_{hv}=40\\ a_{hmax}=50000\\ a_{hmin}=8000}}{}(\mathbf{h})$$

$$\gamma_{K}(\mathbf{h}) = 0.10 + 0.75Exp \underset{\substack{a_{hv}=2\\ a_{hmax}=800\\ a_{hmin}=800}}{}(\mathbf{h}) + 0.15Sph \underset{\substack{a_{hv}=40\\ a_{hmax}=7000\\ a_{hmin}=7000}}{}(\mathbf{h})$$

Compared to the variograms for the normal scores (Figure 6.5), the SC porosity scores exhibit similar continuity as its normal scores counterpart; while the SC permeability shows a very different spatial structure with the variability increasing more quickly at the short scale in both the vertical and the horizontal directions.

Indicator Simulation using Markov-Bayes Model. Unlike Gaussian techniques, the indicator approach is a non-parametric method. The conditional distribution is defined by the kriged estimates for a series of thresholds. This requires variograms for each threshold. The cross-variogram between thresholds should also be modeled and cokriging applied; however, this quickly becomes cumbersome and often impractical as the number of thresholds increases (e.g. for 4 thresholds, 10 variograms are required to be simultaneously modeled for the LMC).

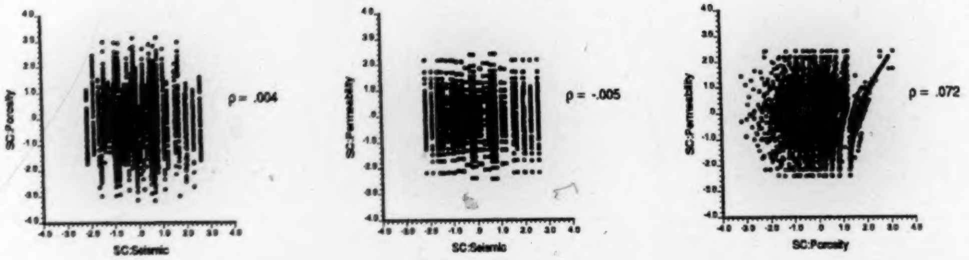


Figure 6.7: Cross plot of stepwise conditionally transformed scores for porosity-permeability (left), porosity-seismic (centre) and permeability-seismic data (right).

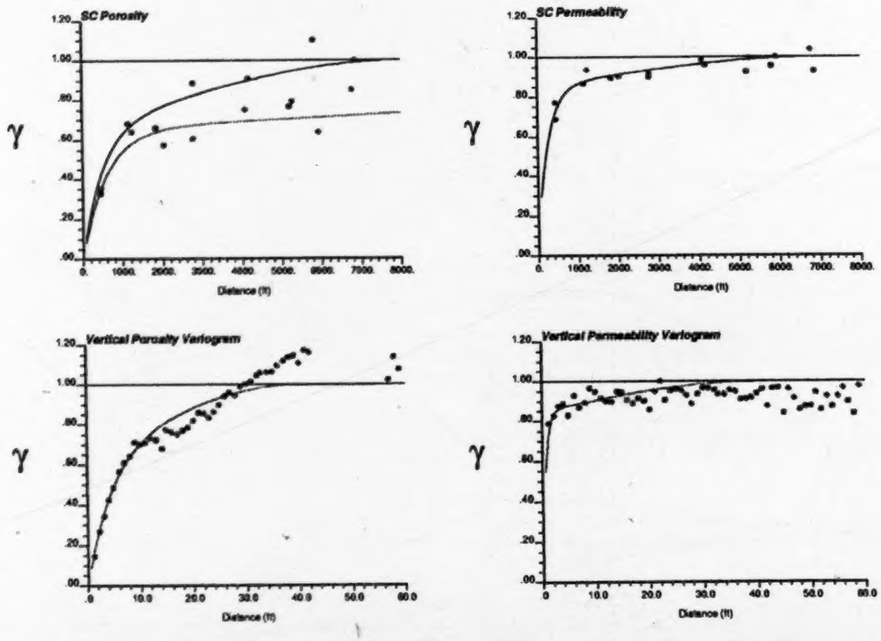


Figure 6.8: Direct variograms for stepwise conditional scores of porosity (left) and permeability (right). Horizontal directions are shown in the top row, while the corresponding vertical variograms are shown in the bottom row. Solid lines are the fitted models for the experimental points shown. Dark lines correspond to the direction of minimum horizontal continuity, while lighter lines correspond to direction of maximum horizontal continuity.

Threshold	$B(\mathbf{u})$	
	Porosity-Seismic	Permeability-Porosity
0.10	0.0510	–
0.20	0.0916	–
0.30	0.1329	0.8814
0.40	0.1433	0.8647
0.50	0.1645	0.8858
0.60	0.1899	0.9038
0.70	0.1534	0.7704
0.80	0.0999	0.6902
0.90	0.0490	0.6198

Table 6.2: Table of calibration factors between the hard local data (left in column title) and available soft data (right in column title).

For this data, porosity was modeled using nine thresholds (corresponding to the 10 deciles). The horizontal and vertical indicator variograms for each threshold are shown in Figures 6.9 and 6.10, respectively. They show similar structures as the Gaussian variograms.

The original permeability data consisted of a large number of 0.05 mD samples, comprising almost 30 per cent. For this reason, permeability was model by defining seven thresholds: 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, and 0.90. The corresponding horizontal and vertical indicator variograms are provided in Figures 6.11 and 6.12, respectively.

To simulate porosity, the gridded seismic data was calibrated to obtain the scaling factors ($B(z)$) to define the cross-covariance between the hard and soft data, as well as the covariance of the soft data (Section 2.1.4). For the simulation of permeability, the porosity data was considered as soft data, and the corresponding calibration factors were also calculated. Table 6.2 lists the calibration factors for the indicator simulation of both porosity and permeability. From Table 6.2, low calibration factors for simulating porosity indicated that the soft seismic data was not as informative than in the case of simulating permeability using soft porosity data.

Comments. Alternative multivariate simulation approaches exist, including principal component analysis (PCA) and direct sequential cosimulation. PCA could be used in multivariate simulation, with a primary objective of reducing the dimension of the required cosimulation (Section 2.2.1). For this particular data set, only two variables are simulated. Depending on the adopted model of coregionalization, there may be relatively little to gain from employing this technique (that is, unless the LMC is chosen, there is little reduced effort in variogram modeling but a rather inconvenient additional transformation and back transformation that must be applied).

Direct sequential cosimulation is another alternative. The theoretical and numerical development is not fully developed. A discussion of the anticipated chal-

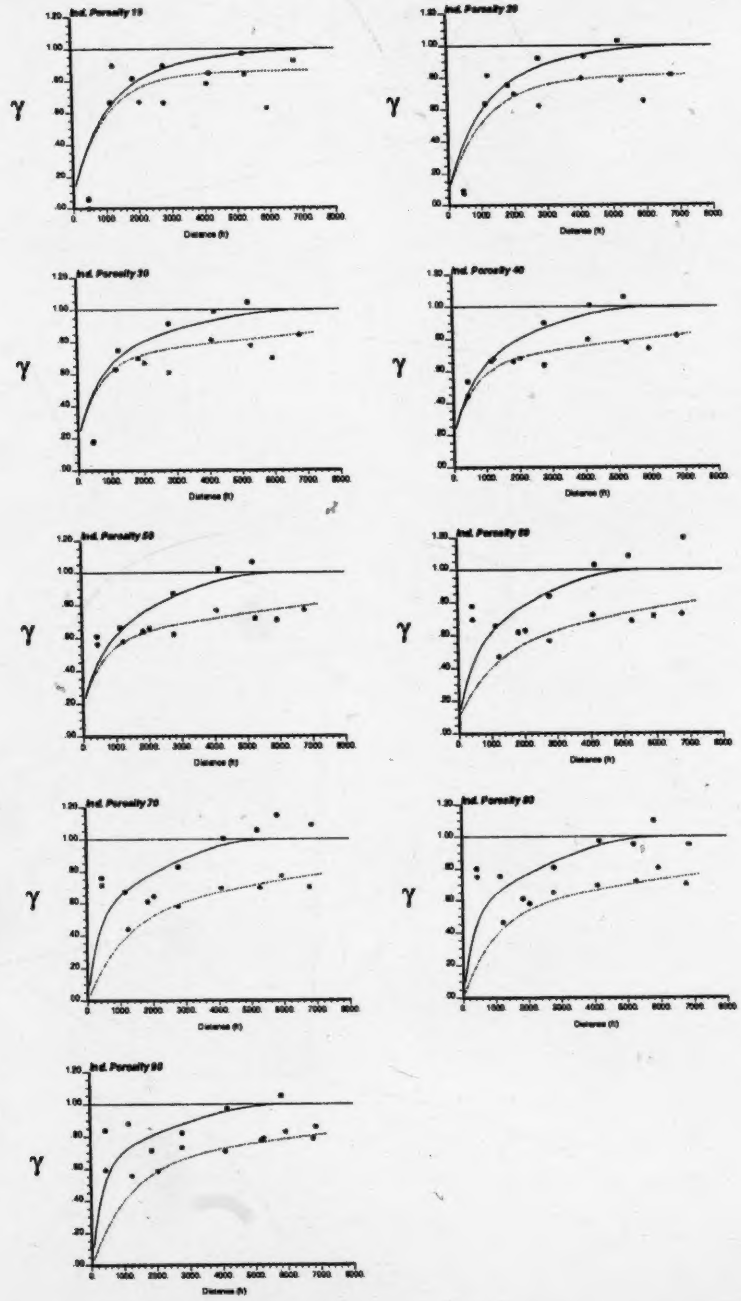


Figure 6.9: Indicator variograms for porosity for each of the nine thresholds. Solid lines are the fitted models for the experimental points shown. Dark lines correspond to the direction of minimum horizontal continuity, while lighter lines correspond to direction of maximum horizontal continuity.

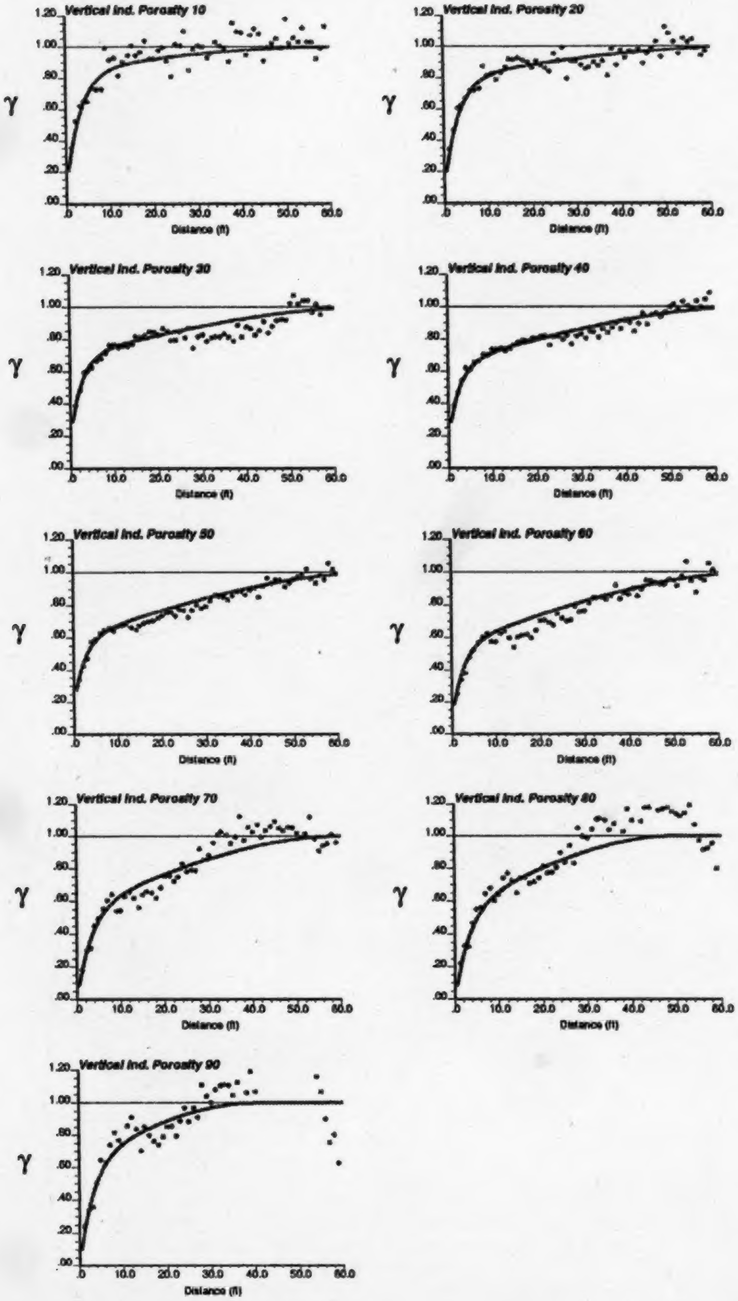


Figure 6.10: Indicator vertical variograms for porosity for each of the nine thresholds. Solid lines are the fitted models for the experimental points shown.

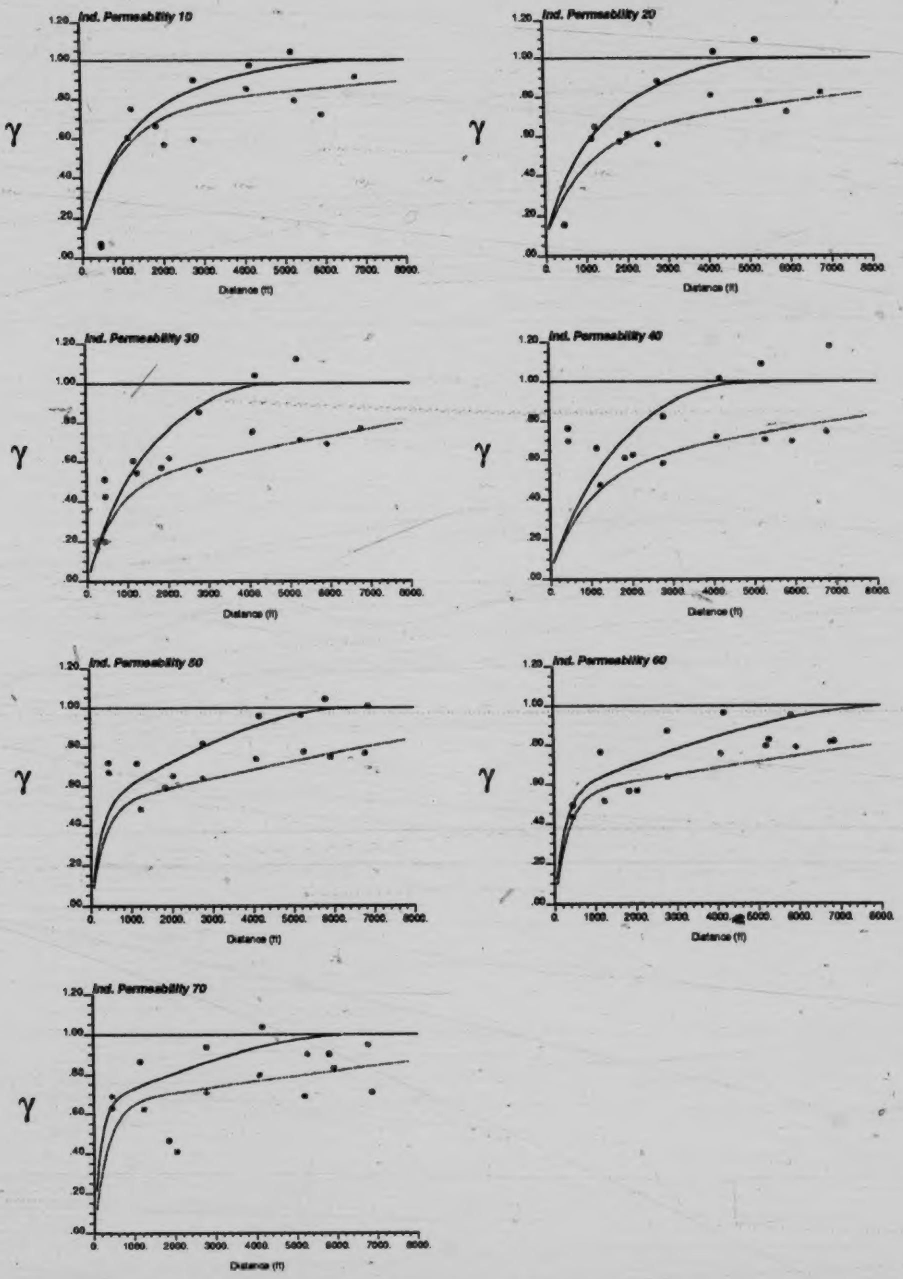


Figure 6.11: Indicator variograms for permeability for each of the seven thresholds. Solid lines are the fitted models for the experimental points shown. Dark lines correspond to the direction of minimum horizontal continuity, while lighter lines correspond to direction of maximum horizontal continuity.

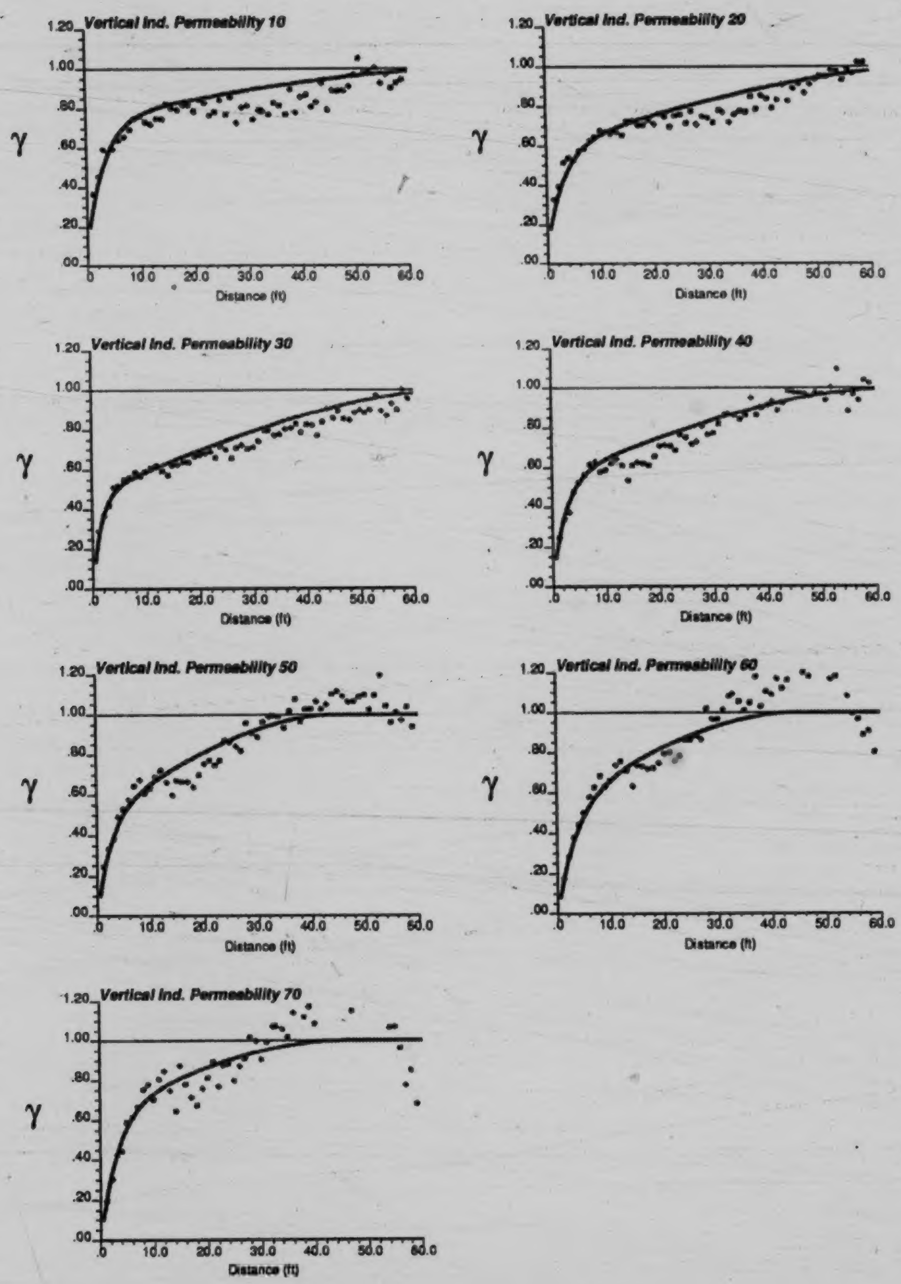


Figure 6.12: Indicator vertical variograms for permeability for each of the seven thresholds. Solid lines are the fitted models for the experimental points shown.

lenges and some ideas being pursued to advance research in this area are provided in Appendix B.

6.3 Comparison of Cosimulation Algorithms

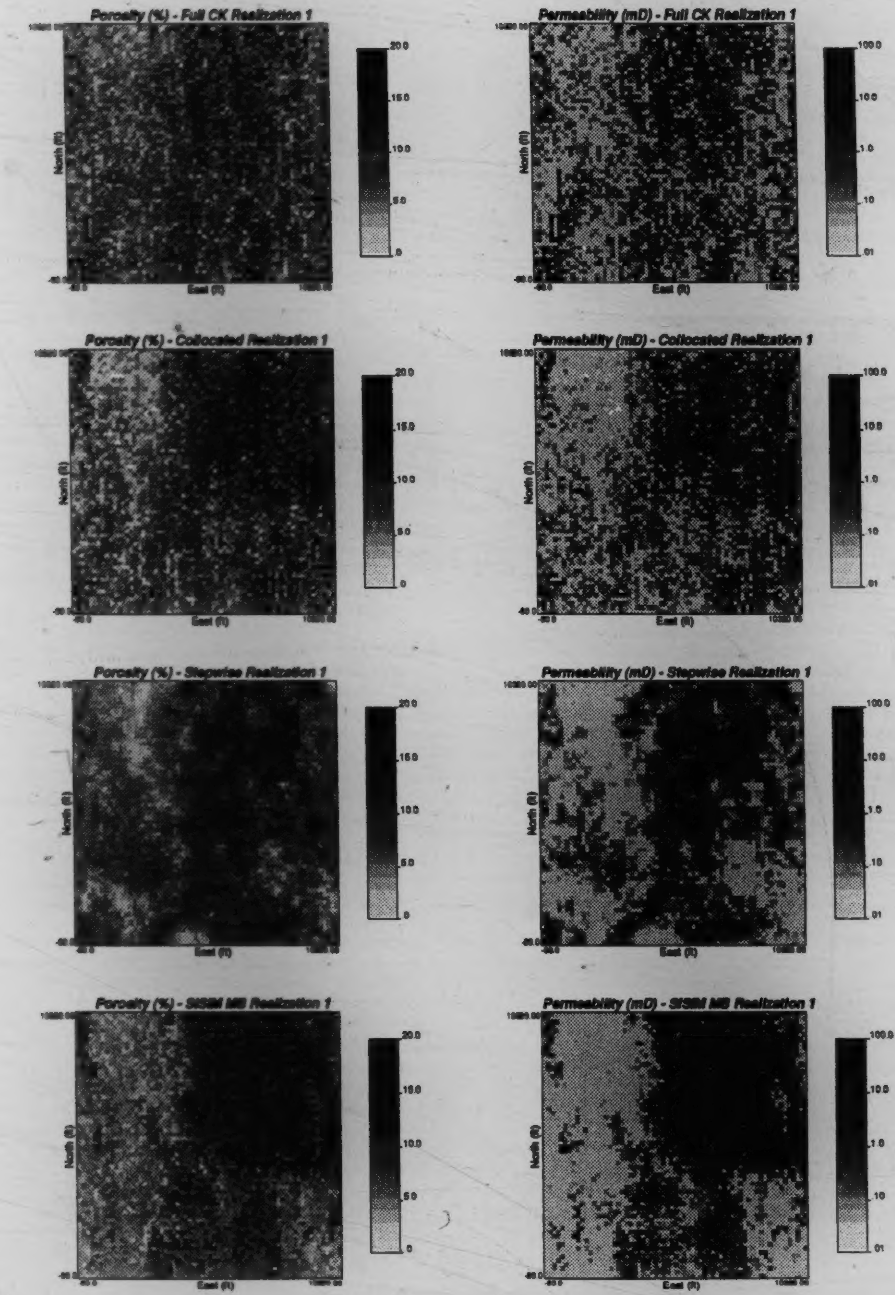
The first realization of porosity and permeability were arbitrarily chosen for the purpose of a simple visual comparison. Figure 6.13 shows the realizations of porosity and permeability, resulting from the four approaches employed. Overall, all four techniques show low values in the west and south-east quadrant, with a high region in the north-east quadrant of the field. These features correspond to those exhibited in the soft data (see Figure 6.1). Of the Gaussian approaches, simulation with stepwise conditionally transformed variables yields the least variable realizations; conventional Gaussian-based simulations all show higher variability. The smoothness of the stepwise conditional realization indicates the strong influence of the seismic data through the multivariate distribution (recall that both porosity and permeability were transformed conditional to seismic data). The Markov-Bayes approach shows the most distinct transition from the low values in the west to the high values in the east.

Aside from visually, it was difficult to compare models generated by each approach without knowledge of the true reservoir. Instead, two comparisons were made: (1) a simple transfer function illustrated the effect on flow performance of the models from each cosimulation method, and (2) an examination of the resulting bivariate distributions showed how well the methods were able to reproduce higher order statistics (extending beyond the traditional histogram and variogram reproduction).

Flow simulation. A simple flow simulation program was used to test the dynamic performance of the realizations from each simulation method. The program `flowsim` by Deutsch [19] provided a quick and simple algorithm that permitted all realizations to be processed. The algorithm involved defining the dimensions of both the input grid and the desired grid (the latter must be some integer, such that the input grid is divisible). Directional effective permeabilities were calculated by setting no flow boundary conditions on parallel sides. For instance, to calculate kx_{eff}, the north and south edges of the grid were set as no flow boundaries, steady-state single phase flow equations were solved to determine the effective permeability in the X direction [19].

Using this flow simulation program, all 100 realizations from each method were processed. Since only an overall comparison was desired, the output grid is set to $1 \times 1 \times 1$ and the north and south edges of the grid were set to no flow boundaries. This provided an overall kx_{eff} for each realization. The methods were then compared based on the $p10$, $p50$ and $p90$ values for kx_{eff} (see Table 6.3 and Figure 6.14).

There was significant spread in the kx_{eff} values. The p_{10} and p_{90} between all the techniques overlap. Gaussian simulation with full cokriging showed the largest spread in values, while indicator simulation with the Markov-Bayes model yields the smallest range. Gaussian simulation with collocated cokriging produces the smallest spread between the p_{10} and p_{90} values of the Gaussian techniques.



Porosity (%) - Full CK Realization 1
Permeability (mD) - Full CK Realization 1
Porosity (%) - Collocated Realization 1
Permeability (mD) - Collocated Realization 1
Porosity (%) - Stepwise Realization 1
Permeability (mD) - Stepwise Realization 1
Porosity (%) - SISIM MB Realization 1
Permeability (mD) - SISIM MB Realization 1
North (ft)
East (ft)
20.0
15.0
10.0
5.0
0
100.0
10.0
1.0
.10
.01

Figure 6.13: First realization of porosity (left) and permeability (right) for each simulation approach: full cokriging (top row), collocated cokriging (second row), stepwise conditional (third row) and indicator with Markov-Bayes (bottom row).

Method	$kx_{eff}(mD)$		
	p_{10}	p_{50}	p_{90}
Full CK	0.1205	0.1965	0.482
Collocated CK	0.167	0.179	0.1925
Stepwise	0.168	0.197	0.233
Markov-Bayes	0.165	0.173	0.18

Table 6.3: Summary of p_{10}, p_{50} and p_{90} effective permeability in X direction from each of the cosimulation methods employed.

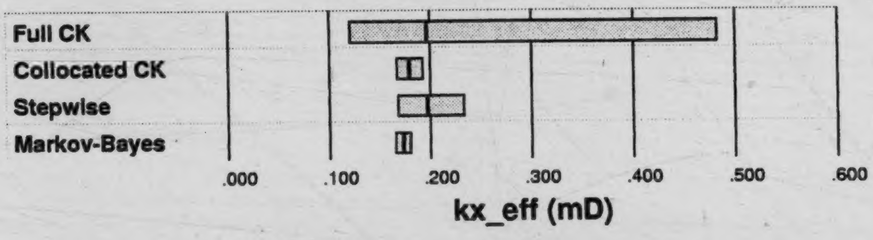


Figure 6.14: Comparison of flow simulation results from different cosimulation algorithms for effective permeability in X direction. The box shows the spread between the $p10$ and the $p90$, while the vertical line within the box shows the $p50$.

Since the true reservoir, and hence its effective permeabilities, were unknown, a comparison of the flow performance of the models generated by each method was not a sufficient measure to compare the performance of the techniques. Instead, we examined the ability of each method to reproduce the original data distributions.

Reproduction of Multivariate Distribution. Recall that the conditioning data was only a small subset of the entire dataset (only mid-well samples were used). We know that simulation reproduces the histogram, and use of simple kriging in simulation will reproduce the variogram. Histograms and variograms were verified, and confirmed these expectations.

Reproduction of the bivariate distribution of porosity and permeability were also examined. These are shown in Figure 6.15 for the first realization. The Gaussian techniques showed good reproduction of the bivariate distribution. The indicator method reproduced the overall shape of the bivariate distribution; however, it produced simulated values with higher variance than those generated by the Gaussian techniques.

Further, a comparison of the bivariate distribution in Gaussian space was compared for the three Gaussian techniques. This required further investigation due to the non-Gaussian features of the normal scores cross plot in Figure 6.4, which corresponds to the transformed data used in the conventional full and collocated cosimulation. Figure 6.16 shows the simulated porosity and permeability values prior to back transformation for the three Gaussian methods. Clearly the full and

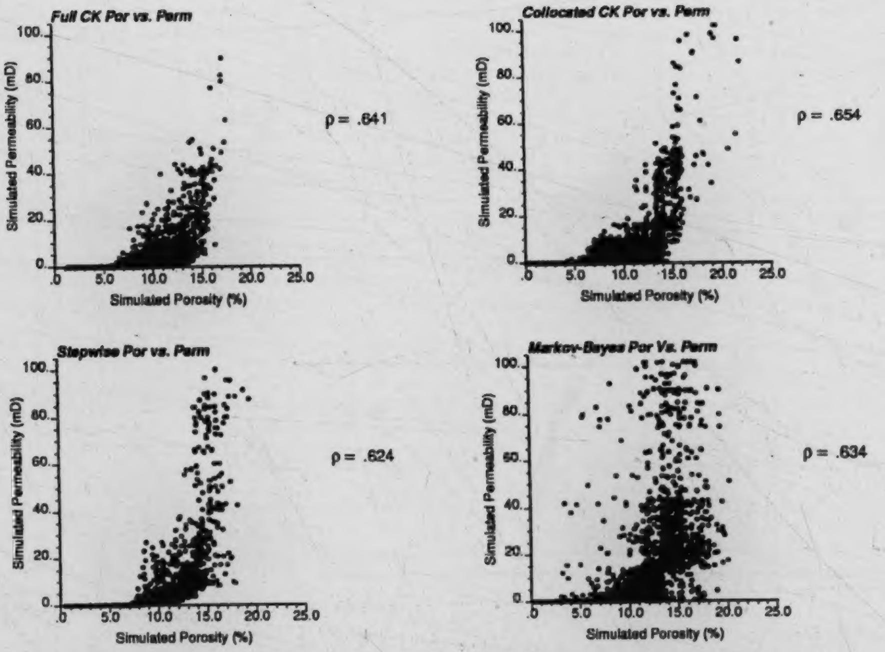


Figure 6.15: Cross plot of porosity and permeability for first realization from each simulation approach: full cokriging (top left), collocated cokriging (top right), stepwise conditional (bottom left) and indicator with Markov-Bayes (bottom right).

collocated cokriging approaches assumed a bivariate Gaussian distribution which was inconsistent with the input bivariate distribution. The stepwise method reproduced the input Gaussian bivariate distribution.

6.4 Remarks

A comparison of multivariate simulation techniques using real data was difficult in the absence of true values. The best we could do was to compare each technique based on their intended application. This required that the assumptions for each approach be considered and compliance with these assumptions be checked.

For this particular dataset, the comparisons of the techniques were based on the ability of the methods to reproduce the multivariate data and compliance with the inherent assumptions of the approach. In the first criteria, simulation of the stepwise conditionally transformed variables showed better reproduction of the bivariate distribution of porosity and permeability. Visualization of the realizations showed the strong influence of the seismic data on the stepwise conditional model. Further, unlike the two conventional Gaussian approaches, it also satisfied the multiGaussian assumptions inherent in the simulation approach.

In practice, the choice of which cosimulation approach to apply will be affected by the amount and type of available data, the number of variables of interest, and the ease of implementation of the technique. This choice will further impact the response variable, and consequently may have significant effects on future decisions pertaining to reservoir development.

In general, the conventional Gaussian cosimulation methods should be applied if the multivariate distribution is approximately multiGaussian after normal score transformation of each variable. The choice between full cokriging and collocated cokriging depends on the number of secondary data that are available and the effort required to infer the LMC model. Collocated cokriging can only be applied if there are secondary data at every location, but variogram modeling is greatly simplified by this approach.

For data that deviate from the Gaussian distribution after normal score transformation, a non-parametric approach may be appropriate. The Markov-Bayes approach would be particularly applicable if the calibration factors indicate that secondary data are informative of the primary variable. Alternatively, if it is important to accurately reproduce distinct features such as non-linearity and/or constraints, then the stepwise approach would be an appropriate choice for simulation.

The advantage of the stepwise approach compared to full cosimulation and Markov-Bayes indicator simulation is the reduced effort in variogram modeling. Relative to both the conventional Gaussian approaches, the stepwise conditional method satisfies the assumption of a multiGaussian distribution inherent in Gaussian simulation. Furthermore, the simplicity of implementation and the computational speed to run a simulation are two large advantages of the stepwise approach. Overall, the stepwise method yields accurate reproduction of the multivariate distribution, and better reproduction of the multivariate features must almost certainly translate to models that are closer to the truth.

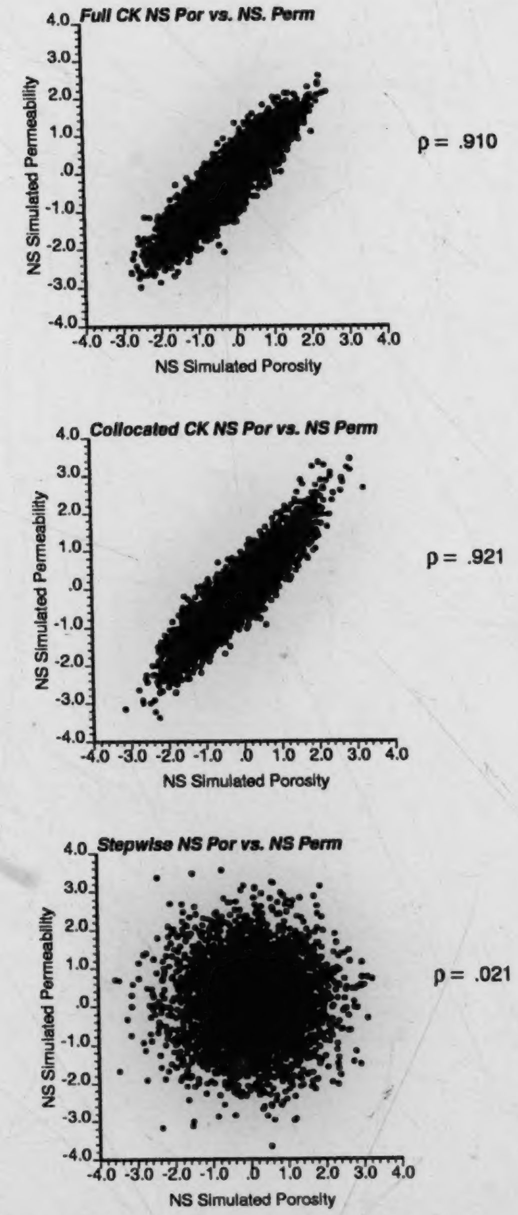


Figure 6.16: Cross plot of normal scores of porosity and permeability for first realization from each simulation approach (prior to back transformation): full cokriging (top), collocated cokriging (middle), and stepwise conditional (bottom).

Chapter 7

Stepwise Transformation for Geostatistical Modeling with a Trend

Trend modeling is an important part of natural resource characterization. A common approach to account for a variable with a trend is to decompose it into a relatively smoothly varying trend and a more variable residual component. Then, the residuals are stochastically modeled independent of the trend. This decomposition can result in values outside the plausible range of variability, such as grades below zero or ratios that exceed 1.0.

In this application of the stepwise conditional transformation, the residuals are transformed conditional to the trend component to explicitly remove these complex features prior to geostatistical modeling. Back transformation of the modeled residual values allows the complex relations to be reproduced.

This chapter discusses some of the methods to detect and model a trend, but will focus primarily on the additive decomposition of the random variable into a mean and residual. Common problems associated to this decomposition will be addressed, and the effectiveness of the stepwise conditional transformation to handle these problems will be shown [47].

7.1 Background

Spatial trends violate the assumption of stationarity (Section 2.1.1) and the application of geostatistical methods is no longer straightforward. Real data often exhibit spatial trends in the first and/or second moment. For example, it is common to have regions of low and high grades within a mineral deposit. Further, the variability within these regions may change depending on the grades. Direct application of common geostatistical tools may inappropriately spread (or smear) spatial features across different areas; trend modeling becomes an integral component to the geostatistical work flow.

A further complication is the subjectivity of trend detection and modeling. There is no "objective" way to determine that there is a trend. The existence of a trend and how to model it is very much dependent on the practitioner. Trends depend on many factors, including the data available and the scale of observation. Although it is common for most trends to be modeled arbitrarily by a decomposition approach, the practitioner's experience with similar deposits/reservoirs may also affect the trend model.

Detecting Trends. In some cases where the depositional environment is well understood, trends can be detected by geological knowledge of the site of interest. In most cases, however, the data are the source for trend detection. Large scale spatial features can be detected during several stages of data analysis and modeling. Sometimes a simple crossplot of the data against elevation may show a trend (Figure 7.1). To visualize trends, a moving window average of the data can be calculated to determine if local means and/or variances are indeed stationary. The size of these "windows" will depend on the number of data available. Also, if few data are available, then these windows may overlap so as to permit more reliable calculation of the local statistics [28, 33]. If there are notable changes in the local mean and variance within the domain, the practitioner may decide that there is a spatial trend.

Although the identification of a trend is subjective, it is widely accepted that the trend is essentially deterministic and should not have short scale variability. Any features that are not significantly larger than the data spacing should probably be left for stochastic modeling.

One further step is to examine the data for a proportional effect, that is, whether the local variance is dependent on the local mean [20, 28, 40]. In general, a crossplot of the local mean and the local variance can show this phenomenon. In the presence of a proportional effect, the relation between the local mean and variance is often quadratic. In this case, application of the stepwise conditional transform involves transforming the residuals to be independent of the mean. This application will often account for the proportional effect; however, some basic checks during model construction can be used to see if further steps are required.

Another stage of the modeling process where spatial trends may be evident is during variography. The experimental variogram may show a trend in any one or more of the principal directions. This is easily identified as the experimental variogram continues to increase above the variance of the random variable as the lag distance, **h**, increases (Figure 7.2). This usually indicates that the practitioner should revisit their decision of stationarity and consider whether the domain should be subdivided or a trend considered.

Common Trend Modeling Approaches. The most common approach is to separate the RV into two components - the trend and the residual:

$$Z(\mathbf{u}) = m(\mathbf{u}) + R(\mathbf{u}) \tag{7.1}$$

where Z is the original RV, m is the trend or mean component, R is the residual RV, and $\mathbf{u}$ denotes the location coordinates (x, y, z). This type of decomposition correspondingly leads to a decomposition of the total variability of the original RV:

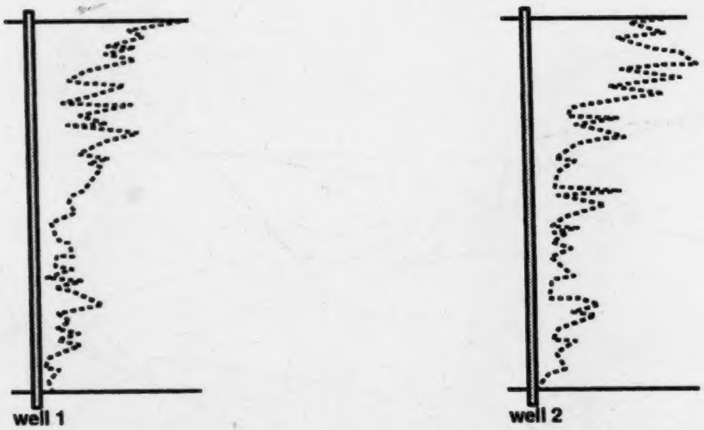


Figure 7.1: Example of vertical trend as indicated by two well logs. (Source: Deutsch, 2002)

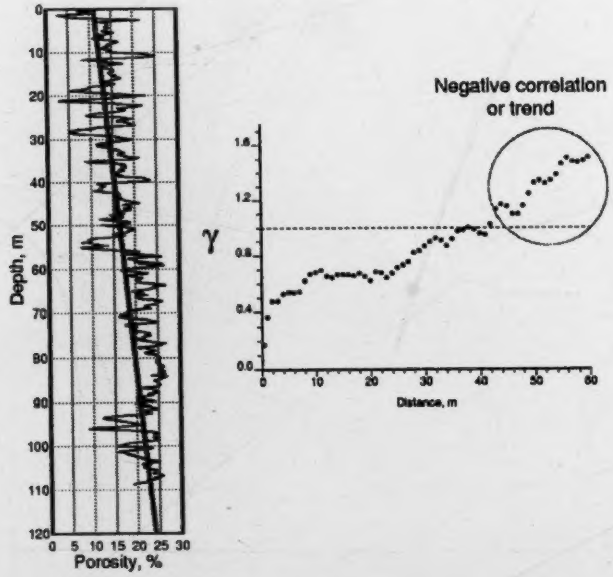


Figure 7.2: Example of porosity log (left) and corresponding vertical variogram (right) showing existence of a vertical trend. (Source: Deutsch, 2002)

$$\sigma_Z^2 = \sigma_m^2 + \sigma_R^2 + 2C(R, m) \tag{7.2}$$

where σ_Z^2 is the variance of the original RV, σ_R^2 is the variance of the residual RV, and $C(R, m)$ is the covariance between the residual and the mean components. This covariance can be either negative or positive; however, if this value is close to zero, fewer artifacts associated to the decomposition are expected [20].

The mean component is defined at all locations via a 3D trend model, while the residual values are only defined at data locations [20]. Geostatistical modeling is then only performed on the residuals, which are considered to be stationary. Multiple realizations of the residuals are generated and added back to the single trend model to produce multiple realizations of the original RV.

The problem remains as to how the trend should be "modeled" so as to obtain a stationary residual random function (RF) for geostatistics. The idea is to obtain a model that accounts for large scale variability; small scale variability is accounted for in geostatistical modeling of the residuals. As a result, trend models are typically smooth models constructed through interpolation *and* extrapolation of the trend data. In areas of interpolation or within the range of the data, there may be no need for a trend model - the model values will be influenced and/or controlled by the data.

There are several trend modeling approaches that have gained popularity in practice, mainly as a result of their ease of application:

1. Hand contour geologic sections to account for drillhole data and analogue information.

2. Calculate moving window averages at each location and use this smooth map as a trend map.

3. Apply common robust estimation algorithms such as ordinary kriging to generate a smooth trend map.

Universal kriging [31] or intrinsic random functions of order k (IRF-k) could also be considered for automatic modeling of the trend. Typically a low order (≤ 2) polynomial function is used to model the trend (a polynomial of order 0 amounts to ordinary kriging with an unknown local mean) [33]. Automatic fitting of the trend using polynomials is generally not recommended as extrapolation of the trend may give rise to unrealistic grades or petrophysical properties. The use of these methods in simulation is problematic and not implemented in most software.

Another common approach to constructing a 3-D trend model is to develop a 1-D and a 2-D trend model and integrate these into a consistent 3-D trend model. A 1-D vertical trend could be developed to capture the trend within drillholes. A 2-D trend map in the horizontal plane could be used to capture any areal trends that may exist between the drillholes. There is no unique way to integrate these two trends into a consistent 3-D trend model [20]; however, one such approach is to scale the areal trend by the proportion of the vertical trend to the global mean:

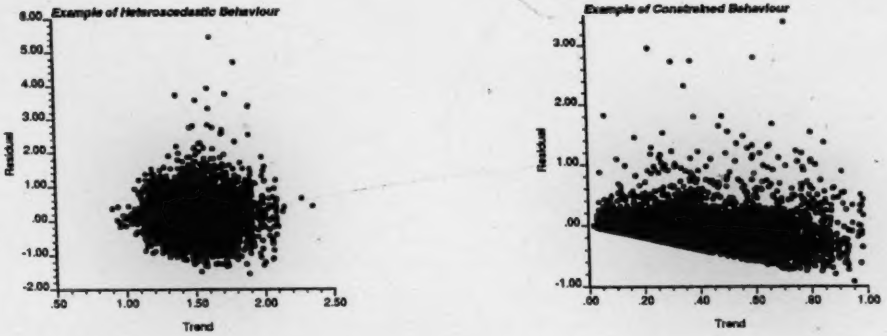


Figure 7.3: Example of heteroscedastic variance of residuals (left), and linear constraint on residuals (right).

$$m(x, y, z) = m_{global} \cdot \left(\frac{m(z)}{m_{global}} \right) \cdot \left(\frac{m(x, y)}{m_{global}} \right) \tag{7.3}$$

This is straightforward and well adapted to practice where limited data may make it difficult to infer a full 3-D trend model. Inherent in Equation 7.3 is an assumption of conditional independence of the vertical trend component within the horizontal plane and the horizontal trend component in the vertical direction.

Problems in Trend Decomposition. Given this common approach of decomposing the RV, the term "trend modeling" has come to be synonymous with the modeling of the local mean. Unfortunately, this is a rather limited view in the sense that trends may exist in both the mean and/or the variance. Common geostatistical estimation and simulation tools, with the exception of indicator approaches, implicitly assume homoscedasticity. Figure 7.3 (left) shows an example of a heteroscedastic relationship between the trend and the residuals. Straightforward application of geostatistical modeling does not account for these departures from stationarity; these must be explicitly handled in the construction of the numerical model of the residual random variable (RV).

The second problem arises as a consequence of the simple decomposition of the RV $Z(\mathbf{u})$ in Equation 7.1. Inevitably, this dissociation results in some constrained bivariate relationship between the trend component, $m(\mathbf{u})$, and the residual component, $R(\mathbf{u})$. For a non-negative RV $Z(\mathbf{u})$, the residual component must be greater than or equal to the negative trend component, that is, $R(\mathbf{u}) \geq -m(\mathbf{u})$. Figure 7.3 (right) shows an example of this type of constraint for a copper deposit for which a 3D trend model was constructed.

The problem arises in the reproduction of this constraint feature after the residuals have been modeled and the trend must be added back to obtain the modeled value of $Z(\mathbf{u})$. A simple addition provides no assurance that $Z(\mathbf{u})$ will be non-negative at unsampled locations.

These two problems of trend modeling must be addressed in order to achieve the initial objectives of constructing numerical models that are geologically realistic and physically plausible.

7.2 Proposed Methodology

The idea is to complement the current practice of trend modeling by applying the stepwise conditional transformation to account for both heteroscedastic and constraint behaviour.

In this application, the residual data are normal score transformed conditional to its trend component. Based on the probability class of the trend component, the corresponding residuals can be conditionally transformed:

$$Y_R(\mathbf{u}) = G^{-1}[F\{R(\mathbf{u}) \mid y_m(\mathbf{u})\}] \tag{7.4}$$

The result is a transformed residual distribution that is standard Gaussian. The transform effectively removes any heteroscedastic or constraint features that may be problematic in the modeling of the residual component.

Much like the forward transformation, the back transformation of the modeled residual values must be conditioned to its collocated trend value. Complex bivariate features are reproduced by way of the back transformation that respects the shape of the multiple conditional distributions.

7.3 Application

7.3.1 Mining Example

The data used in this application was taken from a copper mine. Figure 7.4 shows the location map of the available drillholes alongside the crossplot of Cu grade against elevation, which shows evidence of a vertical trend. The location map in Figure 7.4 also indicates a trend of high values near the centre (and slightly east of the centre) of the map. The trend model was constructed by first calculating a vertical trend. Secondly, a horizontal trend map was generated to give a 2D trend. This involved calculating vertical averages across the horizontal domain from the data. Using these vertical averages, a 2D trend map can be generated by any of the common methods previously mentioned. For this data, the horizontal trend map was created by kriging; Figure 7.5 shows the vertically averaged Cu data that was used as conditioning data in kriging alongside the resulting kriged map.

Regardless of the method chosen to create a 2D trend map, these lower dimension trends must still be integrated into a consistent 3D trend model. Using the 1D and 2D trends shown in Figures 7.4 and 7.5, Equation 7.3 was used to obtain a 3D trend model (see Figure 7.6).

Using the 3D trend model, the residuals were calculated using Equation 7.1. The resulting relation between the trend and the residual was captured in a crossplot shown in Figure 7.7. Clearly, a constraint was imposed on the residual values as a

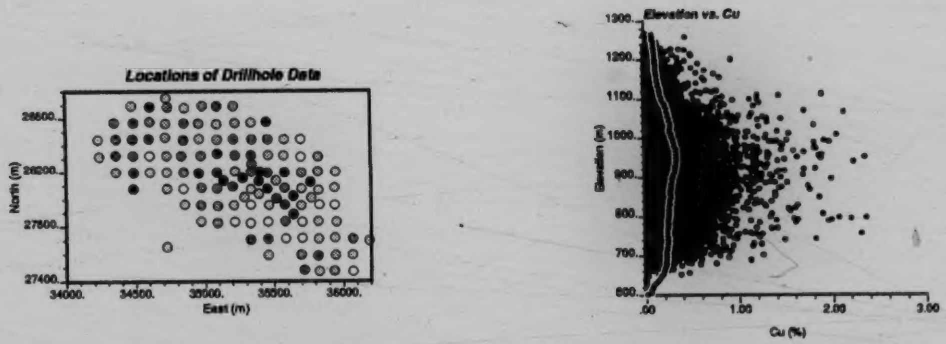


Figure 7.4: Location map of available drillholes (left) and crossplot of elevation vs. Cu to illustrate 1-D trend in the vertical direction (right).

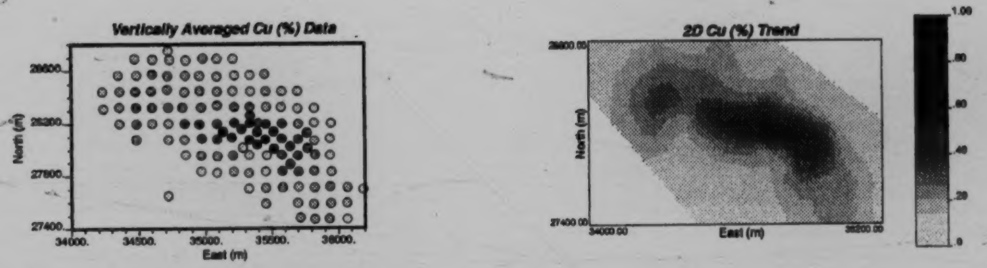

Vertically Averaged Cu (%) Data
North (m)
28600.
28200.
27800.
27400.
34000.
34500.
35000.
35500.
36000.
East (m)
2D Cu (%) Trend
28600.00
27400.00
34000.00
36200.00
North (m)
East (m)
1.00
.80
.60
.40
.20
0

Figure 7.5: Location map of vertically averaged Cu data (left), and resulting kriged map using this data (right).

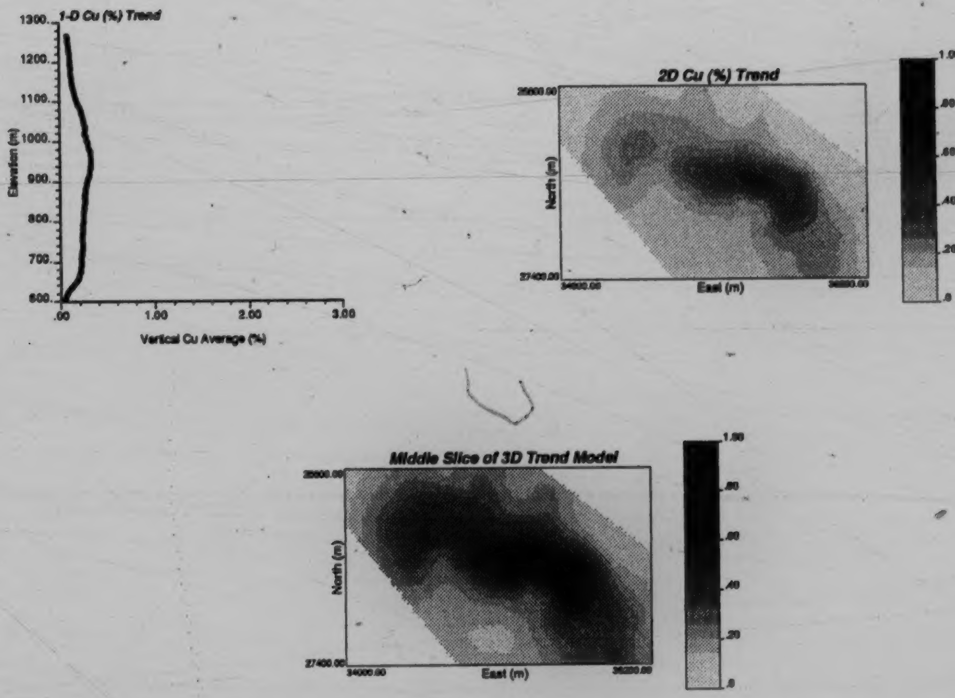


Figure 7.6: Required 1D trend (top left) and 2D trend (top right) for integration to obtain 3D Cu trend model (bottom).

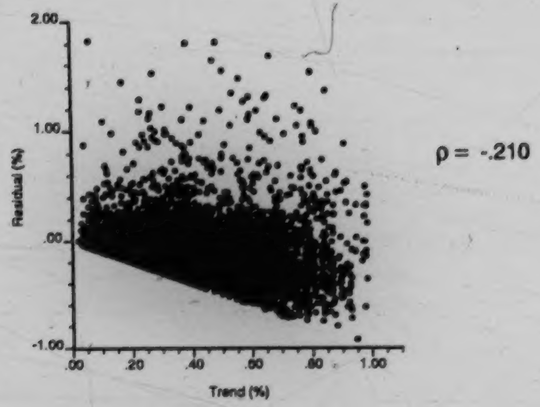


Figure 7.7: Linear constraint on residual Cu values due to trend component.

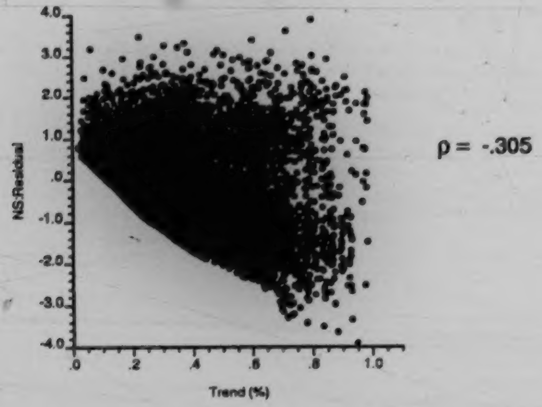


Figure 7.8: Crossplot of Cu trend vs. normal scores of Cu residual.

consequence of the trend model and non-negative grade values. Modeling the residual values to obtain a 3D residual model must reproduce this constraint relationship with the trend in order to obtain non-negative model values of the Cu grade.

To simulate the residuals, sequential Gaussian simulation was used. Applying the conventional normal score transform to the residuals yields the crossplot shown in Figure 7.8. Figure 7.8 clearly shows the transference of the linear constraint in original space (see Figure 7.7) to an almost linear constraint in normal space. The correlation between the mean and transformed residual is -0.305, significant enough to indicate that the two RVs should be modeled in a dependent fashion. Further, the use of popular Gaussian simulation techniques would not be able to reproduce this type of constraint, regardless of whether kriging or cokriging is used.

The stepwise conditional transformation of the residuals was then performed, and the corresponding histograms and crossplot are shown in Figure 7.9. The transformed residuals are univariate Gaussian, and the constraint features have

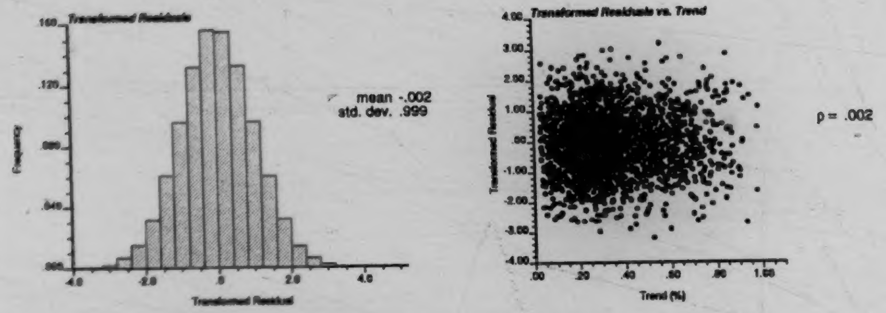


Figure 7.9: Histogram of transformed residual (left) and crossplot of Cu trend vs. the stepwise conditionally transformed residual components (right).

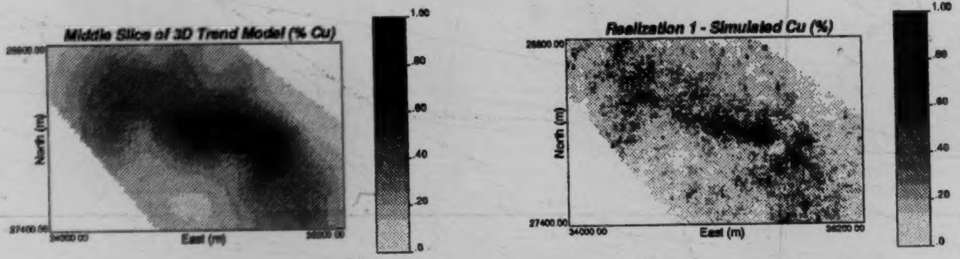


Figure 7.10: Comparison of trend model (left) and simulated realization of Cu (right), after adding the trend back to the simulated residuals.

been removed. Further, the zero correlation combined with homoscedasticity of the resulting bivariate distribution permits independent simulation of the transformed residuals.

Variography and simulation of the conditionally transformed residuals were then performed. Following simulation, the simulated residuals were back transformed. Then, the trend model shown in Figure 7.6 was added to each of the residual realizations to obtain multiple realizations of Cu. One simulated realization of Cu is shown in Figure 7.10.

Figure 7.11 shows the comparison of the distribution of the first realization of simulated Cu and the declustered histogram of the original Cu data. The summary statistics are comparable, as is the shape of the distribution; however, negative values are apparent in the distribution of the simulated Cu.

Negative grades in the simulated Cu values must also be examined. In the first realization, 37 444 of the 817 400 blocks simulated yielded slightly negative Cu grades after the trend and residuals models were added due to imprecision in the classes. This amounted to 4.6 % of the modeled blocks. In comparison, the conventional normal score approach yielded 250 856 negative valued blocks or 30.7 %. The conditional transformation approach provides an obvious improvement from the conventional approach. In fact, all 37 444 negative values fall within the last

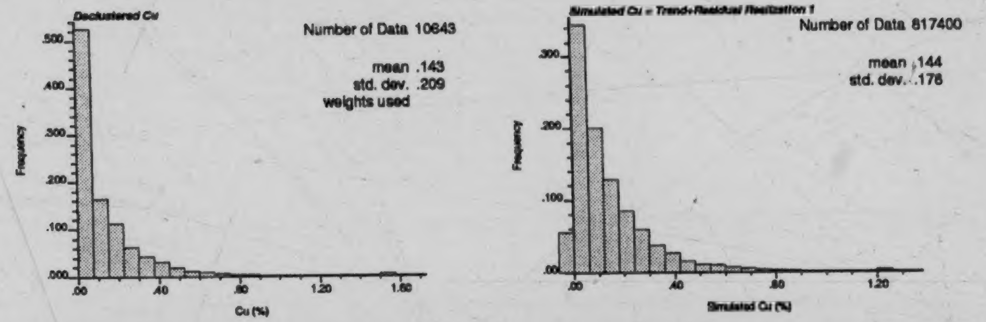


Figure 7.11: Comparison of declustered Cu distribution (left) with the first realization of simulated Cu, after adding the trend component to the simulated residuals (right).

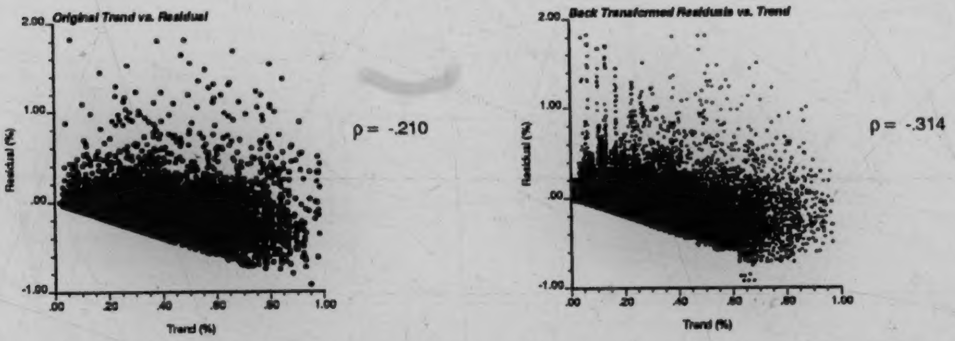


Figure 7.12: Comparison of original Cu trend-residual crossplot and modeled Cu trend - simulated residual crossplot. Notice that the linear constraint from the original crossplot was reproduced.

probability class as specified by a trend value ≥ 0.62. This is consistent with the small group of points in the bottom right corner of Figure 7.12 of the simulated values (right figure).

The real test in this exercise was actually the reproduction of the bivariate relation between the residual and its collocated trend value. This is shown by plotting a single realization of the residuals with the 3D trend model; Figure 7.12 revealed that the linear constraint was reproduced. In contrast, the standard normal score approach produced the crossplot shown in Figure 7.13. Clearly the linear constraint was not reproduced by the conventional transform.

7.3.2 Petroleum Example

Figure 7.14 shows the location map of the available 63 wells and the 1D vertical trend in the well log porosity. Note that for this example, an exaggerated 100:1 stratigraphic coordinate was used.

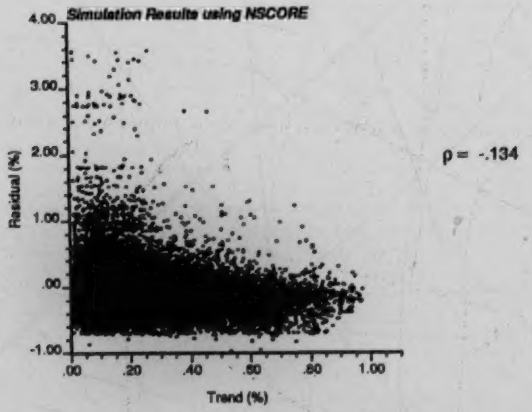


Figure 7.13: Crossplot of the modeled Cu trend vs. simulated Cu residuals from applying the conventional normal score transform. The linear constraint from the original crossplot is not reproduced.

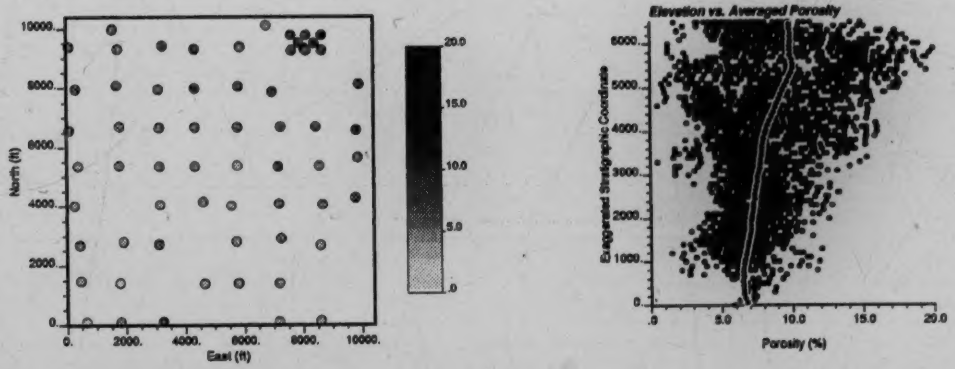


Figure 7.14: Location map of available wells (left) and crossplot of elevation vs. core porosity to illustrate 1-D trend in the vertical direction (right). Note that a 100:1 exaggerated vertical scale was applied.

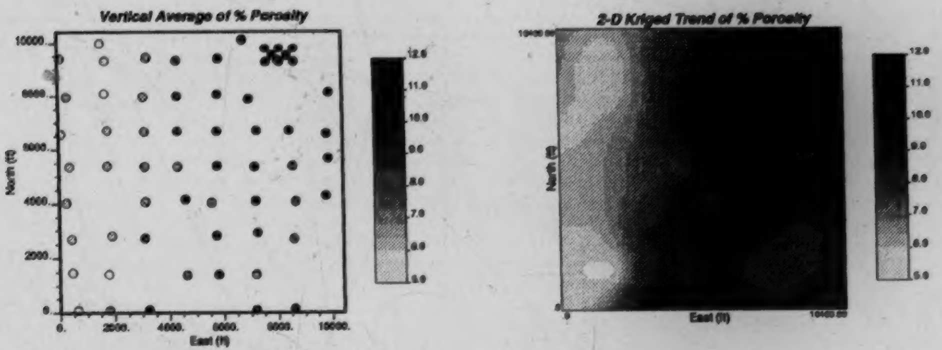


Figure 7.15: Location map of vertically averaged porosity data (left) and resulting areal trend map using this data (right).

Similar to the previous mining example, the data were averaged vertically at each well location to yield 63 conditioning data. These data are input to a 2D kriging to give an areal trend model (Figure 7.15). Equation 7.3 was implemented to integrate the 1D and 2D trends from Figures 7.14 and 7.15 into a consistent 3D trend model (Figure 7.16).

Using this trend model, the residuals were calculated. A crossplot between these residuals and the collocated trend values shows the non-linear relationship in Figure 7.17. Note that although the correlation coefficient is close to zero (0.019), this value only refers to the *linear* relationship and does not adequately reflect any non-linear features.

The stepwise transform was applied and the resulting histogram of the transformed residuals and its relation to the transformed trend component are shown in Figure 7.18. These transformed residuals were then simulated and back transformed. The 3D trend model was then added to the resulting realizations to obtain multiple realizations of porosity. Figure 7.19 provides a comparison of the 3D trend model and one realization of simulated porosity. It shows the reproduction of the large scale features captured by the trend model.

Finally, the histogram of the simulated porosity and the crossplot of the trend and the residual can be compared. Figure 7.20 shows the histogram reproduction of porosity. Figure 7.21 shows the comparison between the crossplot using the available data and the crossplot resulting from the simulated residuals and the trend model. There was good reproduction of both the univariate distribution and the complex bivariate non-linear features.

7.4 Remarks

Note that an alternative application of the stepwise conditional transformation in this particular setting could be performed. Rather than transforming the residuals conditional to the trend, it is possible to transform the actual variable (Cu and porosity in the above examples) conditional to the trend. In fact, the latter trans-

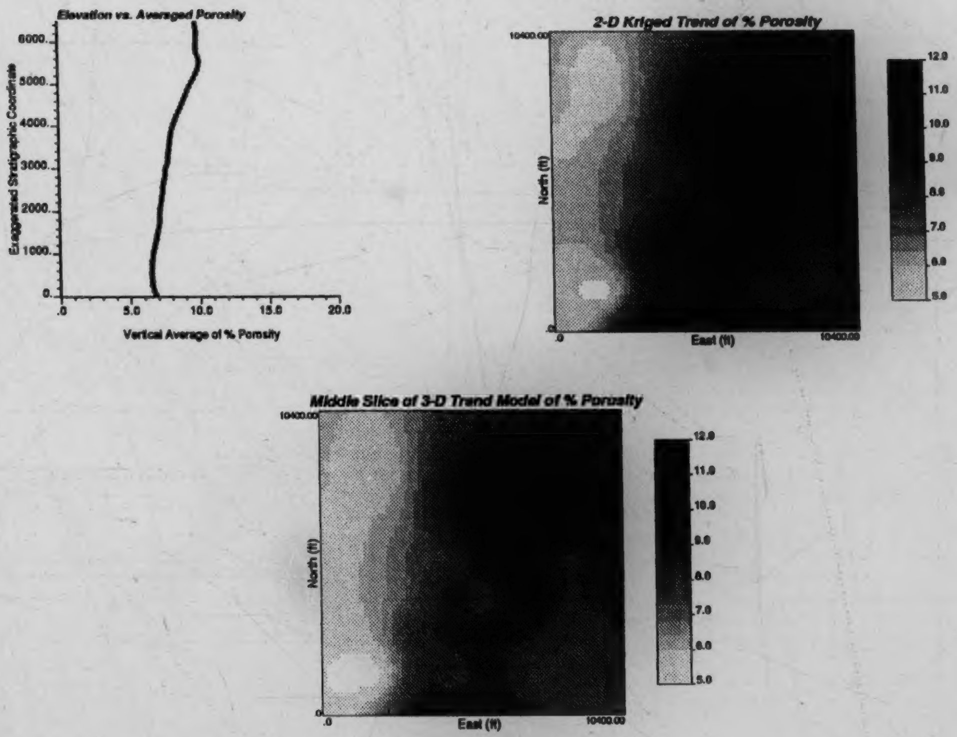


Figure 7.16: Required 1D vertical trend (top left) and 2D areal trend (top right) used to construct a 3D porosity trend (bottom).

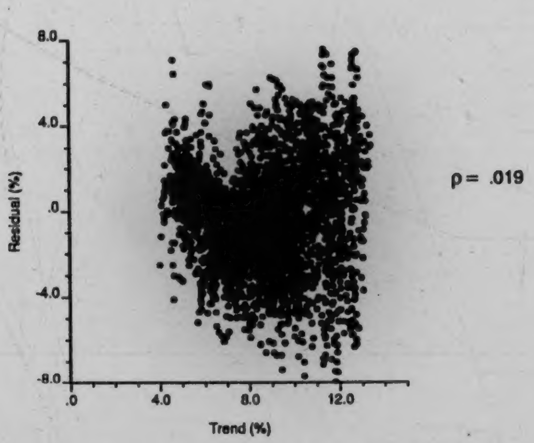


Figure 7.17: Relation between residual porosity values and collocated trend component.

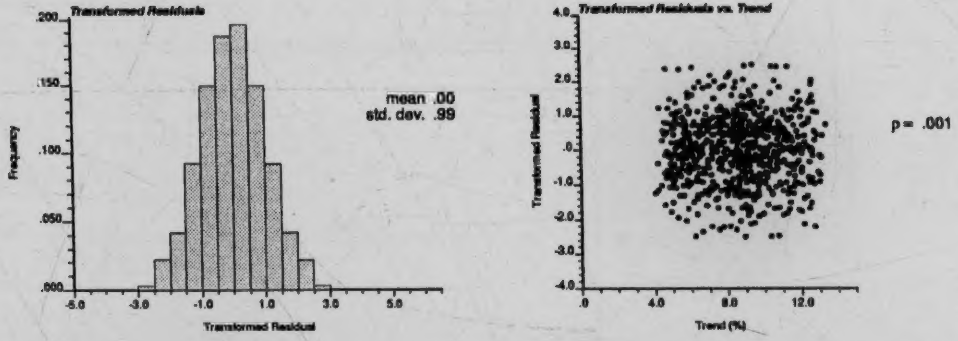


Figure 7.18: Histogram of the stepwise conditionally transformed (SC) residual (left),and crossplot of the trend vs. the SC residual components (right).

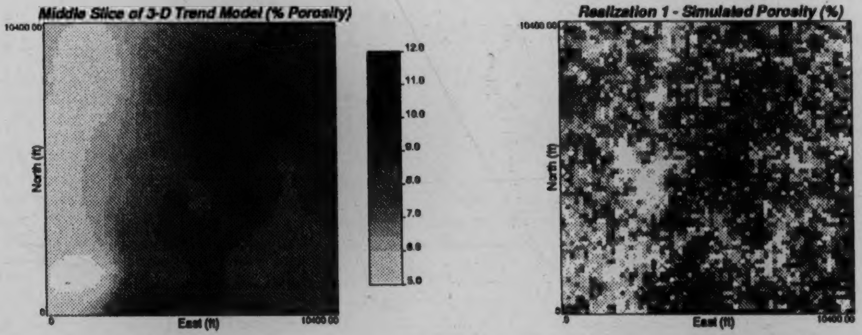

Middle Slice of 3-D Trend Model (% Porosity)
10400.00
0
North (ft)
East (ft)
0
10400.00
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
Realization 1 - Simulated Porosity (%)
10400.00
0
North (ft)
East (ft)
0
10400.00

Figure 7.19: Comparison of porosity trend model (left) and a realization of simulated porosity (right).

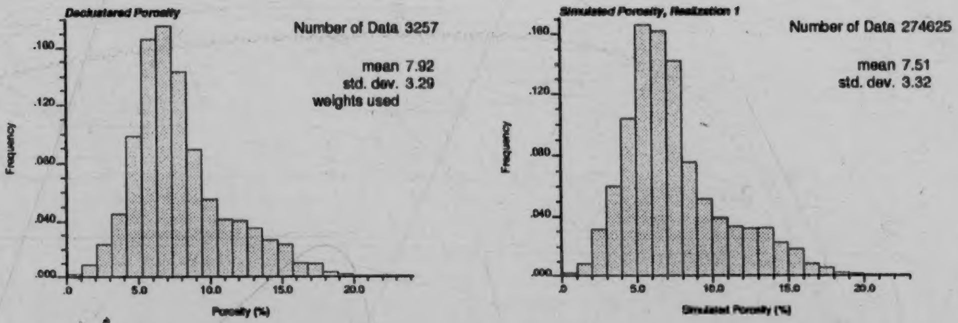


Figure 7.20: Comparison of declustered porosity distribution (left) with the first realization of simulated porosity (right).

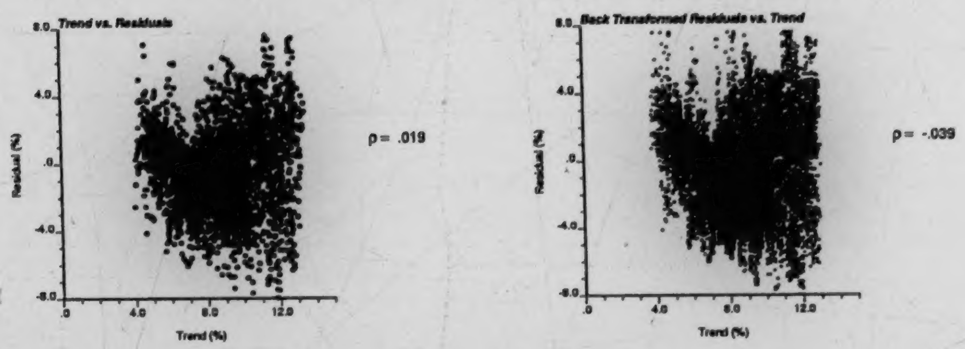


Figure 7.21: Comparison of original porosity trend-residual crossplot and modeled trend - simulated residual crossplot. Notice that the non-linear features were reproduced.

formation applied to Cu in the mining example would have completely removed the potential for negative simulated values. This comes directly from the fact that the original Cu data are all non-negative values. The quantile transformation for any conditional distribution respects the minimum and maximum range of the original data. Since there are no negative Cu values, then back transformation within *any* class must always yield non-negative values.

The stepwise conditional transformation is remarkably robust in its ability to reproduce the complex constraint, non-linear and heteroscedastic relations that may result from conventional trend modeling practices.

Chapter 8

Concluding Remarks

In the case of well behaved multivariate distributions, the conventional approach of independent normal score transformation of each variable may be perfectly adequate in accounting for multivariate features. Unfortunately, the relationship between attributes of interest in naturally occurring phenomena often exhibit mineralogical constraints, heteroscedastic and/or non-linear characteristics. For these complex multivariate relations, the conventional approach does a poor job of reproducing these relations; the stepwise conditional transformation is effective in explicitly accounting for these problematic distributions.

8.1 Stepwise Conditional Transformation

The stepwise conditional transformation is a robust transformation method for multivariate data. Its application as a pre- and post-processing transform to Gaussian simulation, makes it remarkably simple to simulate multiple variables while honouring complex multivariate relations. Several "features" of this technique are important for application to real data:

- No assumption is made on the shape of the input multivariate distribution. The transform removes all structure in the input multivariate distribution, making it particularly robust in handling problematic characteristics of multivariate distributions such as heteroscedasticity, non-linearity and mineralogical-type constraints. Restoration of the input structure is achieved in back transformation.
- The resulting variables are independent at lag distance $\mathbf{h} = 0$ because all conditional distributions are transformed to standard normal distributions.
- Cosimulation may not be required. Independent simulation of the transformed variables can proceed after verification that the cross covariance is approximately zero for all lag distances, i.e. $C'_{ij}(\mathbf{h}) \simeq 0, i \neq j, \mathbf{h} > 0$. Back transformation restores the multivariate dependence between the original variables.
- The covariance structure of the original variables is embedded in the covariance structure of the conditionally transformed variables. This is because the transformed secondary variables are a combination of the original variables.

- The order of the transformation matters since the n^{th} variable is a function of the first $n-1$ variables. Choosing the most continuous variables first works best in practice.

- In presence of insufficient data for reliable inference of all conditional distributions, there are two possible options for application: (1) a smoothing algorithm could be used to "fill in" the multivariate distribution so that reliable conditional distributions can be identified, and (2) allow dynamic class expansion within the transformation so that more data can be used to identify more reliable conditional distributions.

- Non-isotopic sampling (that is, all variables are not available at all locations) is a serious limitation of the technique. If there are some isotopic samples available, then the transform can proceed in one of two ways: (1) choose the more densely sampled variable as the primary variable, or (2) choose the more important variable as the primary and simulate to obtain a value at each location, and then transform the secondary variable conditional to the simulated values of the primary variable.

In practice, multivariate geostatistical simulation typically considers no more than three or four variables for modeling. Although conventional models of coregionalization and the corresponding cosimulation approaches do not theoretically limit the maximum number of variables that can be modeled, the practical demands of implementation to more than four variables makes for a cumbersome and challenging work flow. In this regard, the "nested" application of the stepwise transformation to the Red Dog data showed that the simulation of as many as seven variables is facilitated by the stepwise transform. Moreover, for a reduced area, the simulation approach showed a 3% increase in profit over the conventional estimation approach.

Further, the stepwise conditional transformation is effective not only in the traditional view of simulating multiple variables, but it is also effective in geostatistical modeling with a trend. For this particular application, the transformation allows for the consideration of the trend component collocated with the residual values. This permits the reproduction of any complex relations that may exist between the trend model and the resulting residuals. This reduces the potential for negative simulated grades or ratios that exceed 1.0.

The stepwise conditional transformation is a tool that improves multivariate Gaussian simulation. The resulting models are better in the sense of capturing multivariate relations that are closer to the truth; hence these models are more accurate in reproducing the joint uncertainty. Used in resource planning and management, these models lead to better informed decisions.

8.2 Guidelines for Practical Application

Within the engineering discipline, the impact of any one technique or methodology depends on its widespread use in industry practice. For this reason, some guidelines for applying the stepwise conditional transformation to real data are provided below.

- Choosing the appropriate transform method for Gaussian simulation: Apply the normal score transform, and check the multivariate distribution of the transformed variables. If the multivariate distribution(s) show linear, homoscedastic relations, then apply the conventional normal score transform.

  If the multivariate distributions show any or a combination of non-linear, constraint or heteroscedastic features, then use the stepwise conditional transform.

- If stepwise conditional transform is used, the following decisions must be made:

  1. Choosing the number of variables to transform simultaneously: One consideration is the number of samples that are available. The minimum number of data should be between 10^N to 20^N samples, where N is the number of variables. For instance, for 2 variables, there should be at least 100 samples; for 3 variables, there should be at least 1000 samples. For a large number of variables, transformation of the variables can be performed using pairs or triplets of variables.
  2. Choosing the type of implementation: This depends on the number of data available and the number of variables of interest. If the number of variables is greater than 4, then a nested approach may be warranted given the number of samples available. Otherwise, the conditional transformation of up to three or four variables should not be a problem (given that sufficient data are available).
  3. Choosing the transform order:
     (a) If there are a large number of nonisotopic samples, then choose the most densely sampled variable first; however, if the more sparsely sampled variable is more important and hence its spatial correlation must be preserved, then choose this variable first. In the latter case, estimation can be performed of the more important variable in order to permit transformation of the available secondary variables at other locations.

     (b) If the number of data are approximately equal and isotopic, then a preliminary assessment of the similarities between the structures of the direct variograms for the variables is required. If the direct variograms are similar in type of structure, range and variance contributions, then examine which direct variogram structure yields the closest fit to the experimental cross-variogram structure when it is scaled by the cross correlation coefficient. The variable that gives the best fit should be chosen as the primary variable.

     In the case where the direct variograms for the original variables are dissimilar, the most continuous variable should be chosen as the primary variable. Continuity of a variable can be assessed by comparing the nugget effect, structure type and range of the variogram.

     (c) If exhaustive secondary data variables are available, such as seismic data, then these should be chosen as the primary variable.
  4. Choosing the number of classes: This depends on the number of data available as well as the set of variables to be transformed. A simple

guide is to calculate $C = \sqrt[N]{Ndata}$ where C is the number of classes, $Ndata$ is the number of data available, and N is the number of variables. Too small a choice (≤ 5) may produce transformed variables that are still correlated, while too large a choice may result in poorly defined conditional distributions. Dynamic class expansion works well for the latter case.

5. Choosing the type of smoothing: If sparse data is a concern, then a kernel smoothing algorithm can be applied or a dynamic class expansion can be permitted. If the number of data is low, that is, below the minimum number of available data suggested above, then kernel smoothing should be used. If the concern is sparse data in the transformation of only the last variable, then allow dynamic class expansion.

8.3 Future Research in Multivariate Geostatistics

Geochemical Information for Stepwise Conditional Transformation. Although the stepwise conditional transformation can reproduce mineralogical constraints, it can only do so if there are sufficient data to clearly define the boundaries of the constraint relations. These constraints may represent stoichiometric constraints as different variables compete for space within the rock mass [50]. Lyall (2000) provides an example situation of a rock mass consisting of two variables, X_2Y and XY where the elements X and Y have identical atomic mass [50]. He shows that any rock mass containing a mixture of the two minerals must have ratios of $X : Y$ that must lie between 1 to 2.

Similarly, the consideration of multiple variables with more complex stoichiometric relations could yield a set of non-linear constraints that contains the region of plausible values. No values should be back transformed outside of this envelop. If geochemical information is available and is capable of defining the bounds of these mineralogical envelops, then incorporation of this information into the stepwise conditional transformation should improve the reproduction of these stoichiometric relations.

Stepwise Conditional Transformation for MultiGaussian Kriging. Verly (1984) proposed multiGaussian (MG) kriging as a way to adhere to the homoscedastic, linear assumptions of kriging while estimating the original data variable [72, 74, 75]. His idea was to perform kriging of the normal scores of the data, and then to numerically integrate between the normal scores distribution, defined by the kriged estimate and variance, and the original data distribution, to determine the estimate in original units. In this way, kriging in normal space yields exactly the correct conditional expectation, but in direct space, the mean must be numerically integrated since it does not coincide with the median value.

The stepwise conditional transformation can be used to perform multivariate MG kriging. Rather than krige the normal scores, kriging of the stepwise conditional (SC) scores could be performed. The kriging equations would be no more complicated, only the numerical integration would be based on conditional distributions that must be back transformed in a conditional manner. This conditioning would account for multiple variable estimates, in a similar fashion to cokriging with the exception that multiGaussian assumptions would be respected.

Direct Sequential Simulation for Multiple Variables. Unlike conventional geostatistics, the concept of direct sequential simulation does not involve any requirement for data transformation. The method should reproduce the data distribution and simulation should proceed without any assumptions regarding the input distribution. Furthermore, simulating directly in original space should permit the integration of data of different volume supports. Although the concept is a relatively good one, its application to date has been unsuccessful in presence of non-Gaussian distributions. Recent developments and ideas promise to overcome past limitations [10, 18, 80, 59, 66]. A methodology could be developed to extend direct simulation to include multiple variables: the idea of direct sequential cosimulation.

Recent advances in the inference of univariate conditional distributions [18, 59] provide a key link to the extension of DSS to multivariate multiscale data. Deutsch et. al. (2001) proposed the identification of conditional distributions using normal quantile transformations. The transform between the global histogram in original space and its normal transform is well understood. In Gaussian space, the conditional distributions associated to a certain mean and variance are easy to obtain. The idea then is to perform a reverse quantile transform of the conditional distribution from Gaussian space to find its counterpart in original data space. Using this approach, a database of conditional distributions (in original space) can be prepared. This database can then be used to simulate in direct space without the requirement for data transformation.

Appendix B lays out the basic framework to extend the DSS algorithm to application for multivariate data. It presents a review of the cokriging formalism and a discussion of anticipated theoretical and practical challenges. The proposed methodology for direct sequential cosimulation of multiple multiscale data is also presented in the Appendix. A multivariate distribution scaling approach is proposed to infer the multivariate conditional distribution. Implementation and testing of the proposed methodology is a priority. Future work will involve identification, exploration and validation of other methodologies that may be applied in order to achieve this same objective.

Multivariate Multiple Point Geostatistics. Classical geostatistics relies on two-point geostatistics, that is, reliance on measures like the variogram or the covariance function. The idea behind multiple point geostatistics is to extract more information about the phenomena by accounting for information between more than two data at a time.

Recent work in multiple point geostatistics has focused on the idea of neural networks and training images. In this sense, multiple points accounts for multivariate spatial information.

Accounting for multiple points as well as multiple variables amounts to getting one step closer to defining the spatial law. Currently, the framework for such ambitious goals does not exist, but the notion of a multivariate multiple point geostatistics presents a new and exciting challenge for this still relatively young field.

Bibliography

[1] J. Aitchison. A new approach to null correlations of proportions. *Math. Geology*, 13:175–189, 1981.

[2] J. Aitchison. Logratios and natural laws in compositional data analysis. *Math. Geology*, 31:563–580, 1999.

[3] F. G. Alabert. The practice of fast conditional simulations through the LU decomposition of the covariance matrix. *Mathematical Geology*, 19(5):369–386, 1987.

[4] A. S. Almeida and A. G. Journel. Joint simulation of multiple variables with a Markov-type coregionalization model. *Math Geology*, 26:565–588, 1994.

[5] W. Bashore, U. Araktingi, M. Levy, and W. Schweller. The importance of the geological model for reservoir characterization using geostatistical techniques and the impact of subsequent fluid flow. In *1993 SPE Annual Technical Conference and Exhibition*, Houston, Texas, October 1993. Society of Petroleum Engineers.

[6] H. Beucher, A. Galli, G. Le Loc'h, and C. Ravenne. Including a regional trend in reservoir modelling using the truncated Gaussian method. In A. Soares, editor, *Geostatistics-Troia*, volume 1, pages 555–566. Kluwer, Dordrecht, Holland, 1993.

[7] L. Borgman, M. Taheri, and R. Hagan. Three-dimensional frequency-domain simulations of geological variables. In G. Verly et al., editors, *Geostatistics for natural resources characterization*, volume 2, pages 517–541. Reidel, Dordrecht, Holland, 1984.

[8] G. Bourgault. Using non-gaussian distributions in geostatistical simulations. *Mathematical Geology*, 29:315–334, 1997.

[9] L. Brieman and J. Friedman. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80:580–598, 1985.

[10] J. Caers. Adding local accuracy to direct sequential simulation. *Mathematical Geology*, 32:815–850, 2000.

[11] S. Chapra and R. Canale. *Numerical Methods for Engineers*. McGraw-Hill, Inc., New York, second edition, 1988.

[12] C. Chatfield and A. Collins. *Introduction to Multivariate Analysis*. Chapman and Hall, London, 1980.

[13] J. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons Inc., New York, 1999.

[14] J. Chu, W. Xu, and A. G. Journel. 3-d implementation of geostatistical analyses - the amoco case study. In J. M. Yarus and R. L. Chambers, editors, *Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies*, pages 201–216. AAPG Computer Applications in Geology, No. 3, 1995.

[15] M. David. *Geostatistical Ore Reserve Estimation*. Elsevier, Amsterdam, 1977.

[16] M. W. Davis. Production of conditional simulations via the LU triangular decomposition of the covariance matrix. *Mathematical Geology*, 19(2):91–98, 1987.

[17] C. Deutsch. Geostatistical methods for modelling earth sciences data. Unpublished MIN E 612 Course Notes, University of Alberta, Alberta, Canada, 1999.

[18] C. Deutsch, T. Tran, and Y. Xie. A preliminary report on: An approach to ensure histogram reproduction in direct sequential simulation. Technical report, Centre for Computational Geostatistics, University of Alberta, Edmonton, AB, March 2001.

[19] C. V. Deutsch. Calculating effective absolute permeability in sandstone/shale sequences. *SPE Formation Evaluation*, pages 343–348, September 1989.

[20] C. V. Deutsch. *Geostatistical Reservoir Modeling*. Oxford University Press, New York, 2002.

[21] C. V. Deutsch and A. G. Journel. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 2nd edition, 1998.

[22] P. Dowd. Conditional simulation of inter related beds in an oil deposit. In G. Verly et al., editors, *Geostatistics for natural resources characterization*, volume 2, pages 1031–1043. Reidel, Dordrecht, Holland, 1984.

[23] B. S. Duran and P. L. Odell. *Cluster Analysis (Lecture Notes in Economics and Mathematical Systems,100)*. Springer-Verlag, Berlin, 1974.

[24] B. Everitt. *Cluster Analysis*. Halsted Press, New York, second edition, 1980.

[25] R. Froidevaux. Probability field simulation. In A. Soares, editor, *Geostatistics Troia 1992*, volume 1, pages 73–84. Kluwer, 1993.

[26] R. Gnanadesikan and J. Kettenring. Discriminant analysis and clustering: Panel on discriminant analysis, classification, and clustering. *Statistical Science*, 4:34–69, 1989.

[27] P. Goovaerts. Spatial orthogonality of the principal components computed from coregionalized variables. *Math Geology*, 25:281–302, 1993.

[28] P. Goovaerts. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York, 1997.

[29] P. V. Guoping Xue, Akhil Datta-Gupta and T. Blasingame. Optimal transformations for multiple regression: Application to permeability estimation from well logs. In *1996 Improved Oil Recovery Symposium*, Tulsa, Oklahoma, April 1996. Society of Petroleum Engineers.

[30] J. Harris and H. Stocker. *Handbook of Mathematics and Computational Science*. Springer-Verlag, New York, 1998.

[31] C. J. Huijbregts and G. Matheron. Universal kriging - An optimal approach to trend surface analysis. In *Decision Making in the Mineral Industry*, pages 159–169. Canadian Institute of Mining and Metallurgy, 1971. Special Volume 12.

[32] E. H. Isaaks. *The Application of Monte Carlo Methods to the Analysis of Spatially Correlated Data.* PhD thesis, Stanford University, Stanford, CA, 1990.

[33] E. H. Isaaks and R. M. Srivastava. *An Introduction to Applied Geostatistics.* Oxford University Press, New York, 1989.

[34] A. Izenman. Recent developments in nonparametric density estimation. *Journal of the American Statistical Association*, 86:205–224, 1991.

[35] R. Johnson and D. Wichern. *Applied Multivariate Statistical Analysis.* Prentice Hall, New Jersey, 1998.

[36] A. Journel. Nonparametric estimation of spatial distribution. *Mathematical Geology*, 15:445–468, 1983.

[37] A. Journel. Markov models for cross-covariances. *Mathematical Geology*, 31:955–964, 1999.

[38] A. G. Journel. The deterministic side of geostatistics. *Mathematical Geology*, 17(1):1–15, 1985.

[39] A. G. Journel. *Fundamentals of Geostatistics in Five Lessons.* Volume 8 Short Course in Geology. American Geophysical Union, Washington, D. C., 1989.

[40] A. G. Journel and C. J. Huijbregts. *Mining Geostatistics.* Academic Press, New York, 1978.

[41] A. G. Journel and W. Xu. Posterior identification of histograms conditional to local data. *Math Geology*, 26:323–359, 1994.

[42] A. G. Journel and H. Zhu. Integrating soft seismic data: Markov-Bayes updating, an alternative to cokriging and traditional regression. In *Report 3, Stanford Center for Reservoir Forecasting*, Stanford, CA, May 1990.

[43] D. Lawley and A. Maxwell. *Factor Analysis as a Statistical Method.* Butterworth & Co., London, second edition, 1971.

[44] O. Leuangthong and C. Deutsch. Application of stepwise conditional transformation to improve gaussianity of model variables. Technical report, Centre for Computational Geostatistics, University of Alberta, Edmonton, AB, March 2000.

[45] O. Leuangthong and C. Deutsch. Theoretical insight and practical implementation details of stepwise conditional transformation technique. Technical report, Centre for Computational Geostatistics, University of Alberta, Edmonton, AB, March 2001.

[46] O. Leuangthong and C. Deutsch. Stepwise conditional transformation for simulation of multiple variables. *Mathematical Geology*, 35:155–173, 2003.

[47] O. Leuangthong and C. Deutsch. Transformation of residuals to avoid artifacts in geostatistical modelling with a trend. *Mathematical Geology*, 2003. Submitted.

[48] O. Leuangthong, G. Lyall, and C. Deutsch. Multivariate geostatistical simulation of a nickel laterite deposit. In *2002 Application of Computers & Operations Research in the Mineral Industry (APCOM)*, Phoenix, Arizona, February 2002. Society of Mining Engineers.

[49] G. R. Luster. *Raw Materials for Portland Cement: Applications of Conditional Simulation of Coregionalization.* PhD thesis, Stanford University, Stanford, CA, 1985.

[50] G. Lyall and C. V. Deutsch. Geostatistical modeling of multiple variables in presence of complex trends and mineralogical constraints. In *GEOSTATS 2000*, Cape Town, South Africa, April 2000. Geostatistical Congress.

[51] B. Manly. *Multivariate Statistical Methods.* Chapman and Hall, London, 1986.

[52] J. H. Marbeau and J. E. Marbeau. Formal computation of integrals for 3-D kriging. In M. Armstrong, editor, *Geostatistics*, volume 2, pages 773–784. Kluwer, 1989.

[53] G. Matheron. A simple substitute for conditional expectation: the disjunctive kriging. In Guarascio et al., editors, *Advanced Geostatistics in the Mining Industry*, pages 221–236, Dordrecht, Holland, 1976. Reidel.

[54] J. A. McLennan. The effect of simulation path in sequential gaussian simulation. Technical report, Centre for Computational Geostatistics, University of Alberta, Edmonton, AB, March 2002.

[55] R. A. Olea, editor. *Geostatistical Glossary and Multilingual Dictionary.* Oxford University Press, New York, 1991.

[56] B. Oz and C. Deutsch. Size scaling of cross-correlation between multiple variables. Technical report, Centre for Computational Geostatistics, University of Alberta, Edmonton, AB, March 2001.

[57] B. Oz and C. Deutsch. A short note on the proportional effect and direct sequential simulation. Technical report, Centre for Computational Geostatistics, University of Alberta, Edmonton, AB, March 2002.

[58] B. Oz, C. Deutsch, and P. Frykman. A visual basic program for histogram and variogram scaling. Technical report, Centre for Computational Geostatistics, University of Alberta, Edmonton, AB, March 2000.

[59] B. Oz, C. Deutsch, T. Tran, and Y. Xie. A fortran 90 program for direct sequential simulation with histogram reproduction. *Computers & Geosciences*, page submitted, 2001.

[60] M. Pyrcz and C. Deutsch. Two artifacts of probability field simulation. *Mathematical Geology*, 33:775–800, 2001.

[61] G. Reklaitis, A. Ravindran, and K. Ragsdell. *Engineering Optimization: Methods and Applications.* John Wiley & Sons Inc., New York, 1983.

[62] M. Rosenblatt. Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23(3):470–472, 1952.

[63] D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, New York, 1992.

[64] G. Seber. *Multivariate Observations.* John Wiley and Sons Inc., New York, 1984.

[65] L. Shmaryan and A. Journel. Two markov models and their applications. *Mathematical Geology*, 31:965–988, 1999.

[66] A. Soares. Direct sequential simulation and cosimulation. *Mathematical Geology*, 33:911–926, 2001.

[67] H. Spath. *Cluster Analysis Algorithms for data reduction and classification of objects.* Ellis Horwood Limited, New York, 1980.

[68] R. M. Srivastava. Reservoir characterization with probability field simulation. In *SPE Annual Conference and Exhibition, Washington, DC*, pages 927–938, Washington, DC, October 1992. Society of Petroleum Engineers. SPE paper number24753.

[69] V. Suro-Pérez. Indicator principal component kriging: the multivariate case. In A. Soares, editor, *Geostatistics-Troia*, volume 1, pages 441–454. Kluwer, 1993.

[70] V. Suro-Perez and A. G. Journel. Indicator principal component kriging. *Mathematical Geology*, 23(5):759–788, 1991.

[71] T. T. Tran. Improving variogram reproduction on dense simulation grids. *Computers & Geosciences*, 20(7):1161–1168, 1994.

[72] G. Verly. The multiGaussian approach and its applications to the estimation of local reserves. *Mathematical Geology*, 15(2):259–286, 1983.

[73] G. Verly. MultiGaussian kriging - a complete case study. In R. V. Ramani, editor, *Proceedings of the 19th International APCOM Symposium*, pages 283–298, Littleton, CO, 1986. Society of Mining Engineers.

[74] G. Verly and J. Sullivan. MultiGaussian and probability krigings - application to the Jerritt Canyon deposit. *Mining Engineering*, pages 568–574, June 1985.

[75] G. W. Verly. *Estimation of Spatial Point and Block Distributions: The MultiGaussian Model.* PhD thesis, Stanford University, Stanford, CA, 1984.

[76] H. Wackernagel. *Multivariate Geostatistics.* Springer-Verlag, Berlin, 1995.

[77] T. Wawruch, C. Deutsch, and J. McLennan. Geostatistical analysis of multiple data types that are not available at the same locations. *Mathematical Geology*, 2002. Submitted.

[78] Web. *Teck Cominco Limited 2002 Annual Report.* Website address http://www.teckcominco.com/investors/reports/ar2002/tc-2002-operations.pdf, 2002.

[79] Web. *The Northern Miner website.* Website address http://www.northernminer.com/, 2003.

[80] W. Xu and A. G. Journel. Dssim: A direct sequential simulation algorithm. In *Report 7, Stanford Center for Reservoir Forecasting*, Stanford, CA, May 1994.

[81] W. Xu, T. T. Tran, R. M. Srivastava, and A. G. Journel. Integrating seismic data in reservoir modeling: the collocated cokriging alternative. In *67th Annual Technical Conference and Exhibition*, pages 833–842, Washington, DC, October 1992. Society of Petroleum Engineers. SPE paper # 24742.

[82] G. Xue and A. Datta-Gupta. A new approach to seismic data integration during reservoir characterization using optimal non-parametric transformations. In *1996 SPE Annual Technical Conference and Exhibition*, Denver, Colorado, October 1996. Society of Petroleum Engineers.

[83] H. Zhu and A. G. Journel. Formatting and integrating soft data: Stochastic imaging via the Markov-Bayes algorithm. In A. Soares, editor, *Geostatistics Troia 1992*, volume 1, pages 1–12. Kluwer, 1993.

[84] E. Zwahlen and T. Patzek. A comparison of mapping schemes for reservoir characterization. In *1997 SPE Western Regional Meeting*, pages 147–157, Long Beach, CA, June 1997. Society of Petroleum Engineers. SPE Paper # 38264.

Appendix A

Program Details

Based on the numerous options available for implementing the stepwise conditional transformation, there are two slightly different versions of the program available. All programs were developed in the same format as the Geostatistical Library (GSLIB) [21].

List of Programs:

- **scatsmth_k** performs the kernel smoothing of the bivariate distribution.
- **stepcon** performs the stepwise conditional transformation with all the options discussed in Chapter 4, with the exclusion of the kernel smoothing.
- **stepcon_k** performs the stepwise conditional transformation specifically to read in a transformation table from **scatsmth_k**.
- **backstep** performs the back transformation of the simulated values to original units.

A.1 Kernel Smoothing

The first program, called **scatsmth_k**, is the scatterplot smoothing program that uses the kernel smoothing of Section 4.3. The corresponding parameter file is shown in Figure A.1. As well, the current implementation of this multivariate smoothing is designed for only two variables.

The following is a description of the parameters:

- **datafl**: file with input data to be smoothed. This file must contain the normal scores of the data; the current program is hardcoded to smooth the bivariate distribution in normal space.
- **vcol(1), vcol(2), vcol(3)**: columns for normal scores of variable 1, variable 2 and weights if there are any.
- **tmin, tmax**: trimming limits to filter out data.
- **xscmin, xscmax**: minimum and maximum values for variable 1
- **yscmin, yscmax**: minimum and maximum values for variable 2
- **outfl**: file for output smoothed distribution.
- **transfl**: file with the transformation table to be read into **stepcon_k**.

```
                    Parameters for SCATSMTH_K
                    *************************

START OF PARAMETERS:

../data/cluster.dat          - file with data
4   5  0                     - columns for X, Y, wt
-1.0e21 1.0e21               - trimming limits
-4.0    4.0                  - X min and max
-4.0    4.0                  - Y min and max
scatsmth_k.out               - file for smoothed distribution output
scatsmth_k.trn               - file for transformation table
0.602                        - correlation coefficient
0.05 0.05                    - X and Y variance for kernel density
```

Figure A.1: Parameters for **scatsmth_k**.

- **corr**: correlation of kernel distributions to be populated.
- **usrvarx, usrvary**: user specified variance for variable 1 and 2 for kernel density.

The smoothing in bivariate space is set up like a 2-D grid, consisting of 100 discretizations on both axes. The X and Y minimum and maximum values define the extent of this 2-D grid (in normal space). The output file with the smoothed distribution is reported based on the 100 partitions created for both variables. Each block in the established 100 × 100 grid is reported with its bivariate frequency and its associated Gaussian transform based on the kernel smoothing. The bivariate frequency defines the conditional distribution of the secondary variable given the value of the primary variable.

The transformation table is output from **scatsmth_k** to be read in as the input transformation table in **stepcon_k** (see Figure A.3).

A.2 Stepwise Transformation: Typical Application

The first version of the stepwise conditional transformation, **stepcon**, is used to perform the forward transformation. The corresponding parameters required are shown in Figure A.2 and are explained below:

- **datafl**: file with input data to be transformed.
- **nvart**: number of variables to transform.
- **vcol(i), i=1,...,nvart, iwt**: columns for variables to transform, and column for weights.
- **tmin, tmax**: trimming limits to filter out data.
- **nclass**: number of classes to partition distributions.
- **ipart**: specify type of partitioning: "delta p" refers to a partitioning based on probability thresholds while "delta y" corresponds to a partitioning by equal

```
                  Parameters for STEPCON
                  **********************

START OF PARAMETERS:
../data/cluster.dat     - file with data
3                       - number of variables to transform
13 14 15                - columns for variable transformation
-1.0e21   1.0e21        - trimming limits
10                      - number of classes
1                       - partition by delta p (=1) or delta y (=0)
stepcon.out             - file for output
stepcon.trn             - file for output transformation table
200                     - number of quantiles to report
0                       - allow despiking (1=yes, 0=no)
1                       - allow dynamic class expansion (1=yes, 0=no)
50                      - minimum data per class
0                       - 1=transform according to specified ref dist
histsmth.out            - file with reference dist.
1   2                   - columns for variable and weight
```

Figure A.2: Parameters for **stepcon**.

data value intervals. In most cases, "delta p" is an appropriate choice as this will ensure that the number of data in each class is approximately the same as all other classes for the same variable.

- **outfl**: file for output. This file contains **nvart** columns appended to the original data file with the transformed data values.
- **transfl**: file with the transformation table to be read into **backstep**.
- **nquant**: number of quantiles to report in the tranformation file. In the case of a large dataset, the output transformation file may be large in size so the idea is to allow the user to control the number of values reported to the file.
- **idespike**: option to allow despiking or not. In some cases, despiking may not be desirable. The user is able to specify whether or not the data should be despiked.
- **iexpand**: option to allow dynamic class expansion, 0=no and 1=yes.
- **mindat**: minimum number of data to define a conditional distribution. If **iexpand** is set to 1, then if the minimum number of data is not found, then the class will expand until the minimum number is found.
- **ismooth**: consider a reference distribution for the primary variable, 0=no and 1=yes.
- **smthfl**: if **ismooth**=1, then read this file to get the reference distribution.
- **isvr,iswt**: column for data and weight to define the reference distribution.

```
                    Parameters for STEPCON_K
                    ************************

START OF PARAMETERS:

../data/cluster.dat        - file with data
2                          - number of variables to transform
13 14                      - columns for variable transformation
-1.0e21  1.0e21            - trimming limits
10                         - number of classes
1                          - smoothed distribution, yes=1,no=0
scatsmth_k.trn             - file for input transformation table
stepcon.out                - file for output
stepcon.trn                - file for output transformation table
```

Figure A.3: Parameters for **stepcon_k**.

A.3 Stepwise Transformation: Kernel Smoothing

The stepwise conditional transform program was revised to handle a reference distribution for the secondary variable, where the original versions only applied the reference distribution to the primary variable. This program is called **stepcon_k.par**, and the parameter file **stepcon_k.par** is shown in Figure A.3.

Most parameters are straightforward. The third and fourth lines from the bottom in the parameter file allow use of a bivariate smoothed distribution obtained from **scatsmth_k**. Recall the **scatsmth_k** is currently set to handle only two variables, so the number of variables to transform can only be two.

A.4 Back Transformation

Regardless of whether **stepcon** or **stepcon_k** is applied, the same back transformation program, **backstep** can be used. Figure A.4 shows an example parameter file that is required for the program.

The following is a description of the parameters:

- **nvart**: number of variables to back transform
- **itrans**: for each variable, it is possible to back transform either one or all of the variables
- **datafl**: a file consisting of the data to be back transformed; there must be **nvart** data files.
- **transfl**: file with the transformation table. This can be the transform table from **stepcon** or **scatsmth_k**.
- **outfl**: file for output.
- **ismooth**: flag to identify whether a smoothed bivariate distribution was used in the forward transform.

```
                    Parameters for BACKSTEP
                    ***********************

START OF PARAMETERS:
3                          - number of variables
1  1  1                    - back transform? (0=no, 1=yes)
var1.out                   - file with variable 1
var2.out                   - file with variable 2
var3.out                   - file with variable 3
stepcon.trn                - file with input transformation table
backstep.out               - file for output
0                          - smoothed distribution, 1=yes, 0=no
nspor.trn                  - univariate transformation table for var1
nsper.trn                  - univariate transformation table for var2

If smoothed dist = 0, delete lines for univariate transform table
```

Figure A.4: Parameters for **backstep**.

- **nstransfl**: file with the univariate transformation tables obtained from performing an independent normal score transform on each variable (prior to executing **scatsmth_k**). As before, if **ismooth** is set to 1, then only two variables can be back transformed.

Appendix B

A Framework for Direct Sequential Simulation for Multiple Variables

The idea of direct sequential simulation (DSS) is to simulate in original data units, without assumptions or transformations about the data distribution. This allows the use of multiscale data in DSS.

Recent advances in the inference of univariate conditional distributions [18, 59] provide a key link to the extension of DSS to multivariate multiscale data. Deutsch et. al. proposed the identification of conditional distributions using normal quantile transformations. The transform between the global histogram in original space and its normal transform is well understood. In Gaussian space, the conditional distributions associated to a certain mean and variance are easy to obtain. The idea then is to perform a reverse quantile transform of the conditional distribution from Gaussian space to find its counterpart in original data space. Using this approach, a database of conditional distributions (in original space) can be prepared. This database can then be used to simulate in direct space without the requirement for data transformation.

Thus far, development of the DSS algorithm has been limited to simulating one variable at a time. (Co)kriging provides the mean and variance parameters of conditional univariate distributions. The focus of this research is to simultaneously simulate multiple variables by inferring a conditional *multivariate* distribution. The framework to accomplish this objective is presented, including a review of the cokriging formalism and a discussion of anticipated theoretical and practical challenges. The proposed methodology for direct sequential cosimulation of multiple multiscale data is presented.

Review of Generalized Cokriging

Suppose there are P variables, $Z_p, p = 1, \ldots, P$ with mean μ_p defined on support V_p centered at location $\mathbf{u}_{\alpha p}$, where $\alpha = 1, \ldots, n_p$ and n_p is the number of available data of type p. It is not necessary that the supports $V_p, p = 1, \ldots, P$ be constant.

$$Z_p(\mathbf{u}_{\alpha p}) = \frac{1}{V_p} \int_{V_p} Z_p(\mathbf{u}_{\alpha p}) du$$

The residual of the original Z_p variable about its mean, $Y_p(\mathbf{u}_{\alpha p})$

$$Y_p(\mathbf{u}_{\alpha p}) = Z_p(\mathbf{u}_{\alpha p}) - \mu_p(\mathbf{u}_{\alpha p}), \forall p, \mathbf{u}_{\alpha p}$$

is also defined on support V_p.

Consider estimating $Y_i^*(\mathbf{u})$ as a linear combination of the P data types (where i can be any one of the P variables):

$$Y_i^*(\mathbf{u}) = \sum_{p=1}^{P} \sum_{\alpha}^{n_p} \lambda_{\alpha p} Y_p(\mathbf{u}_{\alpha p})$$

The corresponding estimation variance is

$$\begin{aligned}
\sigma_E^2 &= E\{(Y_i(\mathbf{u}) - Y_i^*(\mathbf{u}))^2\} \\
&= E\{[Y_i(\mathbf{u})]^2 + [Y_i^*(\mathbf{u})]^2 - 2Y_i(\mathbf{u}) \cdot Y_i^*(\mathbf{u})\} \\
&= E\{[Y_i(\mathbf{u})]^2\} + E\{[Y_i^*(\mathbf{u})]^2\} - 2E\{Y_i(\mathbf{u}) \cdot Y_i^*(\mathbf{u})\} \\
&= E\{[Y_i(\mathbf{u})]^2\} + E\left\{\sum_{p=1}^{P} \sum_{p'=1}^{P} \sum_{\alpha=1}^{n_p} \sum_{\beta=1}^{n_{p'}} \lambda_{\alpha p} \lambda_{\beta p'} Y_p(\mathbf{u}_{\alpha p}) Y_{p'}(\mathbf{u}_{\beta p'})\right\} \\
&\quad -2E\left\{\sum_{p=1}^{P} \sum_{\alpha}^{n_p} \lambda_p(\mathbf{u}_{\alpha p}) Y_i(\mathbf{u}) Y_p(\mathbf{u}_{\alpha p})\right\} \\
&= E\{[Y_i(\mathbf{u})]^2\} + \sum_{p=1}^{P} \sum_{p'=1}^{P} \sum_{\alpha=1}^{n_p} \sum_{\beta=1}^{n_{p'}} \lambda_{\alpha p} \lambda_{\beta p'} E\{Y_p(\mathbf{u}_{\alpha p}) Y_{p'}(\mathbf{u}_{\beta p'})\} \\
&\quad -2 \sum_{p=1}^{P} \sum_{\alpha}^{n_p} \lambda_p(\mathbf{u}_{\alpha p}) E\{Y_i(\mathbf{u}) Y_p(\mathbf{u}_{\alpha p})\} \\
&= \bar{C}(V_i(\mathbf{u}), V_i(\mathbf{u})) + \sum_{p=1}^{P} \sum_{p'=1}^{P} \sum_{\alpha=1}^{n_p} \sum_{\beta=1}^{n_{p'}} \lambda_{\alpha p} \lambda_{\beta p'} \bar{C}(V_p(\mathbf{u}_{\alpha p}), V_{p'}(\mathbf{u}_{\beta p'})) \\
&\quad -2 \sum_{p=1}^{P} \sum_{\alpha}^{n_p} \lambda_p(\mathbf{u}_{\alpha p}) \bar{C}(V_i(\mathbf{u}), V_p(\mathbf{u}_{\alpha p}))
\end{aligned}$$

with

$$\bar{C}(V_p(\mathbf{u}_{\alpha p}), V_{p'}(\mathbf{u}_{\beta p'})) = \frac{1}{V_p \cdot V_{p'}} \int_{V_p} du \int_{V_{p'}} C(V_p(\mathbf{u}_{\alpha p}), V_{p'}(\mathbf{u}_{\beta p'})) dv$$

Minimizing the error variance with respect to the weights gives the $\sum_{p=1}^{P} n_p$ equations that constitute the simple cokriging system of equations:

$$\sum_{p'=1}^{P} \sum_{\beta=1}^{n_{p'}} \lambda_{\beta p'} \bar{C}(V_p(\mathbf{u}_{\alpha p}), V_{p'}(\mathbf{u}_{\beta p'})) = \bar{C}(V_i(\mathbf{u}), V_p(\mathbf{u}_{\alpha p})) \qquad \text{(B.1)}$$

where $p = 1, \ldots, P$ and $\alpha = 1, \ldots, n_p$. In matrix notation, the left hand side covariance matrix consists of $P \times P$ submatrices of volume to volume covariances between different data types.

$$[[\bar{\mathbf{C}}(\mathbf{V_p}, \mathbf{V_{p'}})], p, p' = 1, \ldots, P]$$

where each submatrix consists of $n_p \times n_{p'}$ covariances between the p and the p' data:

$$\bar{\mathbf{C}}(\mathbf{V_p}, \mathbf{V_{p'}}) = \begin{bmatrix} \bar{C}(V_p(\mathbf{u}_{1p}), V_{p'}(\mathbf{u}_{1p'})) & \cdots & \bar{C}(V_p(\mathbf{u}_{1p}), V_{p'}(\mathbf{u}_{n_{p}p'})) \\ \vdots & \ddots & \vdots \\ \bar{C}(V_p(\mathbf{u}_{n_pp}), V_{p'}(\mathbf{u}_{1p'})) & \cdots & \bar{C}(V_p(\mathbf{u}_{n_pp}), V_{p'}(\mathbf{u}_{n_pp'})) \end{bmatrix}$$

The large covariance matrix (containing all submatrices) is symmetric.

$$[\bar{\mathbf{C}}(\mathbf{V_p}, \mathbf{V_{p'}})] = [\bar{\mathbf{C}}(\mathbf{V_{p'}}, \mathbf{V_p})]^T$$

The column vector of weights and right hand side covariances then consists of $\sum_{p=1}^{P} n_p$ elements:

$$\lambda = \begin{bmatrix} \lambda_{11} \\ \vdots \\ \lambda_{n_1 1} \\ \\ \vdots \\ \vdots \\ \\ \lambda_{1P} \\ \vdots \\ \lambda_{n_P P} \end{bmatrix} \qquad \bar{\mathbf{C}}(\mathbf{V_i}(\mathbf{u}), \mathbf{V_p}(\mathbf{u}_{\alpha\mathbf{p}})) = \begin{bmatrix} \bar{C}(V_i(\mathbf{u}), V_1(\mathbf{u}_{11})) \\ \vdots \\ \bar{C}(V_i(\mathbf{u}), V_1(\mathbf{u}_{n_1 1})) \\ \\ \vdots \\ \vdots \\ \\ \bar{C}(V_i(\mathbf{u}), V_P(\mathbf{u}_{1P})) \\ \vdots \\ \bar{C}(V_i(\mathbf{u}), V_P(\mathbf{u}_{n_P P})) \end{bmatrix}$$

Solving for the weights in the cokriging system of equations (B.1) gives the minimized error variance known as the simple cokriging (SCK) variance

$$\sigma^2_{SCK} = \bar{C}(V_i(\mathbf{u}), V_i(\mathbf{u})) - \sum_{p=1}^{P} \sum_{\alpha=1}^{n_p} \lambda_{\alpha p} \bar{C}(V_i(\mathbf{u}), V_p(\mathbf{u}_{\alpha p}))$$

The resulting cokriging estimate and estimation variance correspond to the conditional expectation and variance of the RV $Y_i(\mathbf{u})$. The above cokriging system corresponds to simple kriging; however, it is straightforward to modify the above formalism to reflect the unit sum constraint of the weights for ordinary kriging.

Simultaneous Cokriging of Multiscale Data. We can consider the simultaneous cokriging of M multiple data types simply by changing the column vector of weights and right hand side covariance into an $M \times P$ matrix, where $M \leq P$.

$$Y_1^*(\mathbf{u}) = \sum_{p=1}^{P} \sum_{\alpha}^{n_p} \lambda^1_{\alpha p} Y_p(\mathbf{u}_{\alpha p})$$

$$\vdots$$

$$Y_M^*(\mathbf{u}) = \sum_{p=1}^{P} \sum_{\alpha}^{n_p} \lambda^M_{\alpha p} Y_p(\mathbf{u}_{\alpha p})$$

Solving for the weights of the resulting cokriging system requires little additional effort since the large left hand side covariance only has to be inverted once. Matrix multiplication of the inverted covariance matrix with the additional $M-1$ columns of the right hand side covariance will give the weights to estimate the other $M-1$ additional variables.

$$\lambda = \begin{bmatrix} \lambda_{11}^{1} & \cdots & \lambda_{11}^{M} \\ \vdots & & \vdots \\ \lambda_{n_1 1}^{1} & \cdots & \lambda_{n_1 1}^{M} \\ & & \\ \vdots & & \vdots \\ \vdots & & \vdots \\ & & \\ \lambda_{1P}^{1} & \cdots & \lambda_{1P}^{M} \\ \vdots & & \vdots \\ \lambda_{n_P P}^{1} & \cdots & \lambda_{n_P P}^{M} \end{bmatrix}$$

$$\bar{\mathbf{C}}(\mathbf{V_1}(\mathbf{u}), \mathbf{V_p}(\mathbf{u_{\alpha p}})) = \begin{bmatrix} \bar{C}(V_1(\mathbf{u}), V_1(\mathbf{u}_{11})) & \cdots & \bar{C}(V_M(\mathbf{u}), V_1(\mathbf{u}_{11})) \\ \vdots & & \vdots \\ \bar{C}(V_1(\mathbf{u}), V_1(\mathbf{u}_{n_1 1})) & \cdots & \bar{C}(V_M(\mathbf{u}), V_1(\mathbf{u}_{n_1 1})) \\ & & \\ \vdots & & \vdots \\ \vdots & & \vdots \\ & & \\ \bar{C}(V_1(\mathbf{u}), V_P(\mathbf{u}_{1P})) & \cdots & \bar{C}(V_M(\mathbf{u}), V_P(\mathbf{u}_{1P})) \\ \vdots & & \vdots \\ \bar{C}(V_1(\mathbf{u}), V_P(\mathbf{u}_{n_P P})) & \cdots & \bar{C}(V_M(\mathbf{u}), V_P(\mathbf{u}_{n_P P})) \end{bmatrix}$$

The only additional computations required in order to simultaneously estimate the collocated data types is the determination of the right hand side volume to volume covariance between the location to be estimated and the nearby data of P types.

While cokriging of one variable gives the conditional expectation and variance of the RV, simultaneous cokriging of multiple RVs gives the conditional mean vector and covariance matrix of the M RVs. Simulation using these distributional parameters must still be performed.

Example of Cokriging of Multiscale Data

Consider two types of data, 7 core samples of variable Y_1 and 3 seismic data of variable Y_2. The core data are considered to be point-scale data, and the seismic data are block scale data informing a 50×50 volume. Suppose we are interested in cokriging an intermediate 10×10 volume. Without loss of generality, suppose both variables, Y_1 and Y_2, have the same direct isotropic variogram:

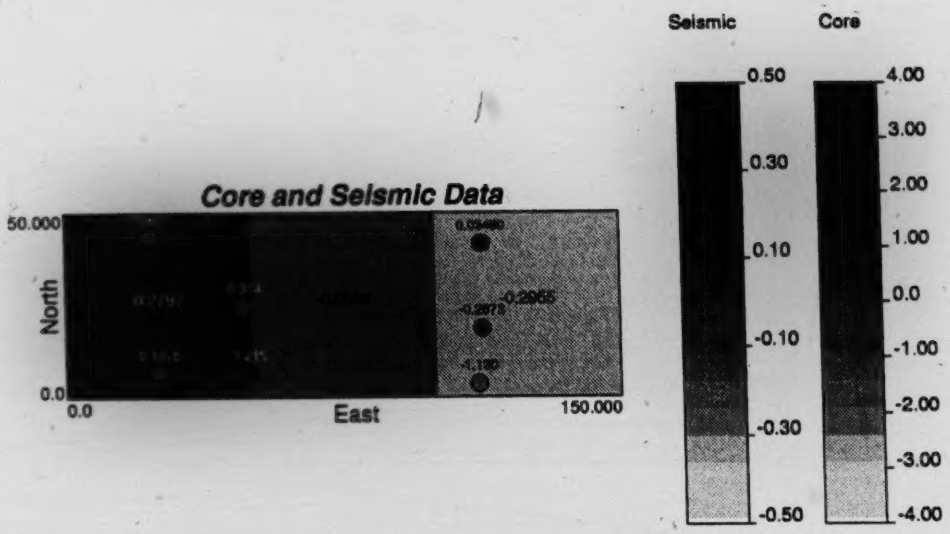


Figure B.1: 2-D map of core and seismic data. Seismic data values are shown in larger font centered on the 50 × 50 volume which it informs.

$$\gamma(\mathbf{h}) = 0.5Sph_{a=3}(\mathbf{h}) + 0.5Sph_{a=15}(\mathbf{h})$$

The correlation between Y_1 and Y_2 was chosen to be 0.70, with an intrinsic cross variogram:

$$\gamma(\mathbf{h}) \quad = \quad 0.35Sph_{a=3}(\mathbf{h}) + 0.35Sph_{a=15}(\mathbf{h})$$

In practice, we calculate these point-scale statistics by downscaling the seismic statistics (see Oz et. al. [58, 56]). We create a consistent data set by simulation.

Two unconditional Gaussian simulations were used to generate two reference 2-D maps at the point scale on a 150 × 50 grid. The first map is considered the reference map of the core data. Seven samples were drawn from this map to give the 7 core samples that will be used for cokriging. The second map was generated by cokriging with a correlation of 0.70. This "reference" map was then upscaled to the 50 × 50 volume to provide the 3 seismic data. Figure B.1 shows both the core and seismic data on the same map.

Cokriging is performed by setting up the simple cokriging equations (see Equation B.1). Average volume to volume covariances are numerically calculated by discretizing each volume into 25 points (i.e. 5 × 5 discretizations), calculating the covariance values between the points in one volume and the points in the second volume, and then averaging these covariance values. To set up the left hand side covariance matrix (i.e. covariance between the data and themselves), the direct variograms were used to get the average covariance between two data of the same type, while the cross variogram is used to determine the average covariance between two different types of data. Similarly, the right hand covariance vector (i.e. covariance between the data and the volume to be estimated) relies on the direct variogram for data of the same type and the cross variogram for data of different types. The key in this latter calculation is that the intermediate scale (which is the volume that we are interested in) is the same type of "data" as the core data and not the seismic data. This means that the average covariance between the core data and the volume to be estimated uses the direct variogram, while the average covariance between the seismic data and this same volume uses the cross variogram. Note that in all cases, it is the point scale variograms that are used in the numerical calculation.

For the 10 × 10 volume centered at coordinates (35,35), the resulting cokriged estimate is 0.15882 with an estimation variance of 0.41576. This simple exercise for one variable can easily be extended to the simultaneous cokriging of multiple variables (as discussed above) to obtain the conditional mean vector and covariance matrix. This multiscale cokriging is key to DSS for multiple variables; however, some theoretical and practical challenges still exist in the implementation of a direct sequential cosimulation (DSCosim) algorithm.

Theoretical and Practical Challenges of DSS for Multiple Variables

Classical geostatistical simulation relies on both kriging and Monte Carlo simulation to obtain simulated values. DSS is no different. Kriging (or cokriging in the case of multiple variables) is used to determine the mean and variance of the conditional univariate distribution at the location to be simulated. Assuming that the two distributional parameters (mean and variance) fully define the conditional distribution, Monte Carlo simulation is performed to obtain a simulated value.

For multivariate geostatistics using multiscale data, inference of the coregionalization model is still a challenge. In the case of multiscale data, downscaling the spatial measures of the larger scale data to the same scale as the smaller scale data is especially challenging. It requires the inference of a short scale structure for a volume that is smaller than that informed by the data. In the previous cokriging example, the average covariance values were numerically determined using the point scale variograms. In the case of real data where the point scale variogram of the large scale data (such as seismic) is unknown, programs exist to downscale the block scale variogram to a point scale (or some small finite scale) variogram; however, it is the inference of the coregionalization model at the point scale that is most challenging. This requires:

1. Inference of a same-scaled cross-variogram based on different scaled data;
2. Downscaling of the direct block-scale variogram to a direct point-scale variogram; and
3. Iterating between Steps 1 and 2 to ensure (i) a legitimate model of coregionalization, and (ii) consistency between the model variograms and the variability of the natural phenomena.

The inference of the coregionalization model is a difficult issue in any conventional multivariate geostatistics that accounts for multiscale data.

Two other challenges that are a result of simulating in direct space are (1) the limitations of using kriging in simulation, and (2) inference of the multivariate distribution.

Implications due to Kriging

Kriging is a linear estimator. The kriging estimate is also the conditional expectation of the RV given the conditioning data. A consequence of linearity in the conditional expectation is the inability to reproduce complex non-linear features. Unfortunately, real data exhibit complex relations.

Kriging also provides information about the uncertainty in its estimate. This is the kriging variance. The variance is independent of the data values and the estimate, hence it is homoscedastic. In contrast, the variance of mineral grades or petrophysical properties found in a real deposit or reservoir is heteroscedastic. For example, it is common to find a low variance in a low valued area, and a correspondingly high variance in a high valued area. The use of the kriging variance does not account for heteroscedastic behaviour of the conditional distribution.

Inference of Multivariate Distribution

Simultaneous kriging of all variables yields the conditional mean vector and the conditional variance corresponding to each variable. All that is required are the correlation coefficients at $\mathbf{h} = 0$ in order to fully define the covariance matrix of the multivariate distribution at $\mathbf{h} = 0$. Given that these correlations are known or can be estimated (as in the case of non-isotopic sampling [77]), and that the multivariate distribution is fully defined by its mean vector and covariance matrix, simulation can proceed by recursive application of Bayes relation.

In the conventional Gaussian framework, the mean vector and covariance matrix provides all the information required to define the multivariate distribution. However, in direct space this information is insufficient. In fact, multivariate distributions of real data generally do not follow nice parametric forms. In these instances, knowing only the mean vector and covariance matrix is not sufficient to define the multivariate distribution.

Deutsch et. al. proposed a methodology to determine the conditional univariate distributions with only the mean and variance provided from kriging [18, 59]. Using this approach, the conditional marginal distribution of each variable can be obtained. The main challenge is then to identify the conditional multivariate distribution knowing these conditional *marginal* distributions.

Updating Technique to Obtain Conditional Multivariate Distribution

Let's review the information that is readily available: the original data distributions consisting of the global (or standard) univariate and multivariate distributions, and the conditional (or non-standard) univariate distributions obtained from solving the cokriging system and the algorithm presented by Deutsch et. al. [18, 59]. To determine a non-standard multivariate distribution, a simple iterative updating procedure using all the available distributions is proposed.

Consider the bivariate case, where we have Y_1 and Y_2 data. The algorithm proceeds as follows:

1. Update the global bivariate distribution, $f_{Y_1,Y_2}(y_1,y_2)$, by scaling it by the ratios of the non-standard univariate conditional distribution, $f_{Y_i'}(y_i)$, to the standard univariate distribution, $f_{Y_i}(y_i)$, $i = 1,\ldots,2$.

$$f_{Y_1',Y_2'}(y_1,y_2) = f_{Y_1,Y_2}(y_2,y_1) \cdot \frac{f_{Y_1'}(y_1)}{f_{Y_1}(y_1)} \cdot \frac{f_{Y_2'}(y_2)}{f_{Y_2}(y_2)} \qquad \text{(B.2)}$$

2. Calculate the new marginal Y_1 and Y_2 distributions that result from Equation B.2 and reset $f_{Y_i}(y_i)$, $i = 1,\ldots,2$ to these new marginals:

$$f_{Y_1}(y_1) = \int_{Y_2} f_{Y_1',Y_2'}(y_1,y_2)dy_2$$
$$f_{Y_2}(y_2) = \int_{Y_1} f_{Y_1',Y_2'}(y_1,y_2)dy_1$$

Also, reset $f_{Y_1,Y_2}(y_1,y_2)$ to the new updated global distribution of Equation B.2:

$$f_{Y_1,Y_2}(y_1,y_2) = f_{Y_1',Y_2'}(y_1,y_2)$$

3. Calculate corresponding summary statistics: mean, variance of marginal distributions, and covariance and correlation of (updated) bivariate distribution.

4. Check that the mean and variance of new marginal distributions $f_{Y_1}(y_1)$ and $f_{Y_2}(y_2)$ match those of the desired conditional distributions, $f_{Y_1'}(y_1)$ and $f_{Y_2'}(y_2)$, within some acceptable margin, ϵ. If this condition is not met, go to Step 1.

Scaling of the multivariate distribution to obtain a conditional multivariate distribution should reproduce complex, non-linear and/or heteroscedastic properties. Note that spatial heteroscedasticity of the simulated values (as mentioned in the section on kriging implications) is different from heteroscedasticity in the multivariate distribution at $\mathbf{h} = 0$ (for which this updating process will account).

Validation of Updating Approach

Consider a simple bivariate Gaussian global distribution with correlation ρ. Given parameters that specify a conditional distribution, the above updating methodology can be applied. For this purpose, a numerical exercise was devised for which the following parameters were set:

- Global Distribution is standard bivariate Gaussian with correlation of $\rho_{global} = 0.30$.
- Conditional Distribution parameters:

$$\begin{aligned} E\{X\} &= 1.2 \\ Var\{X\} &= 0.25 \\ E\{Y\} &= 0.0 \\ Var\{Y\} &= 1.0 \end{aligned}$$

Furthermore, we know that restandardizing the marginal X and Y distributions to non-standard means and variances will not change the correlation of the resulting bivariate distribution.

The updating approach was applied and the corresponding results are shown in Figure B.2. The reference global and desired conditional distributions are shown at the top (left and right, respectively). Three iterations were required to obtain an updated conditional distribution that reproduced the desired conditional univariate statistics; however, the resulting updated conditional distribution has a correlation of 0.161, not the 0.30 that would be found by rescaling a bivariate Gaussian distribution.

Unfortunately, re-standardizing the marginal distribution of X to some other non-standard Gaussian distribution and then calculating the resulting bivariate distribution does not amount to a conditional distribution in the conventional sense of the term. Rather, this generates a new global bivariate distribution with the desired univariate marginal distributions. Thus, the reference conditional distribution may be incorrect, that is, the expected correlation of 0.30 may not be correct. To investigate the difference between the correlation of the conditional distribution with that resulting from the iterative scaling approach, a slight shift in the thought process is proposed.

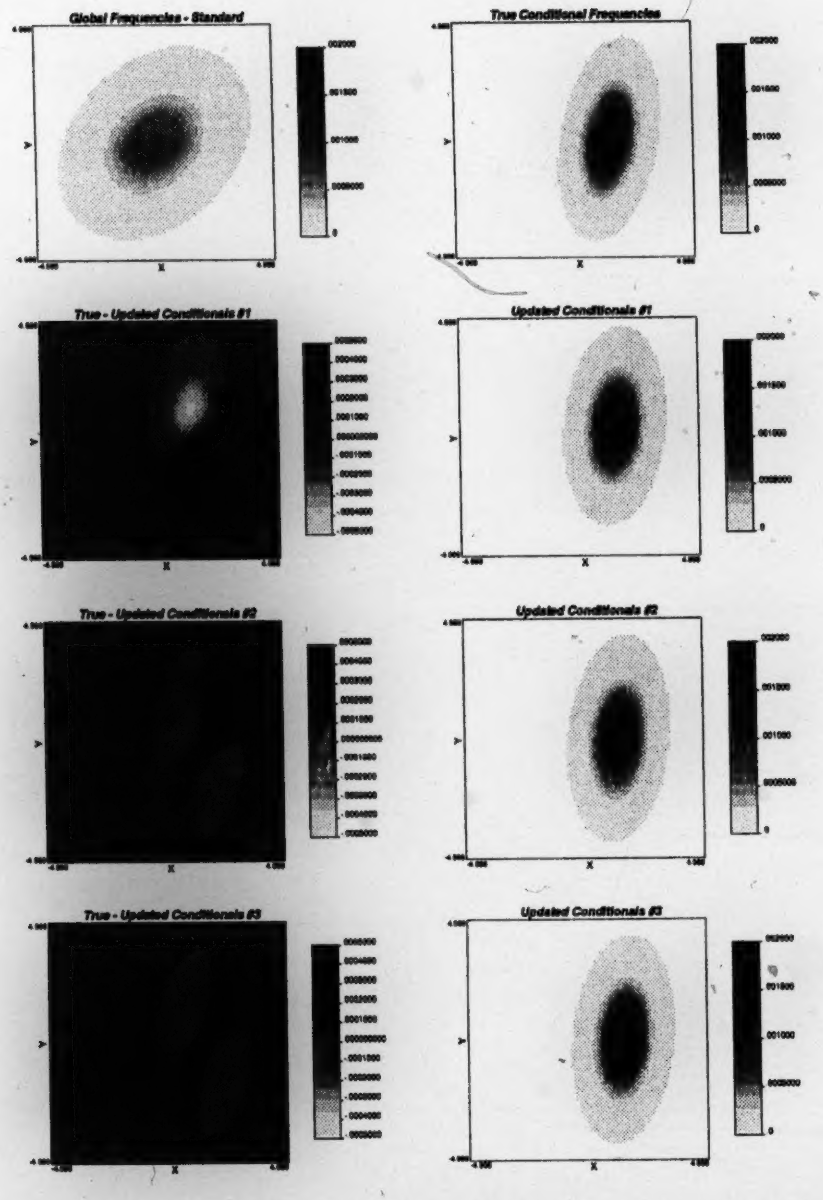


Figure B.2: Results from applying iterative updating approach to a global standard bivariate Gaussian distribution with correlation of 0.30 (top right). Reference conditional univariate distribution for X is set with a mean of 1.20 and a variance of 0.25 (top left). Difference in bivariate probability distribution between reference conditional distribution and corresponding updated conditional distribution are shown in the remaining left side plots. The updated conditional distributions (for each iteration) are shown in the bottom 3 right side plots.

Building a Global Distribution from Conditional Distributions

Suppose that the global multivariate distribution is a linear combination of conditional multivariate distributions. This is analogous to supposing that these conditional distributions are subsets of the multivariate distribution, resulting from considering only a subset of the data in the domain. This idea is consistent with the practical application of (co)kriging, which only considers a subset of the data that are within some neighbourhood of the location of interest.

Without loss of generality, consider that there are m bivariate Gaussian conditional distributions, all with common correlation, ρ. The resulting global bivariate distribution is a linear combination of these m distributions (and will be non-Gaussian unless $m \rightarrow \infty$):

$$f(x,y) = \sum_{i=1}^{m} p_i \cdot f_i(x,y)$$

where $p_i, i = 1,\ldots,m$ are weights corresponding to $f_i(x,y), i = 1,\ldots,m$. The weights represent the contribution of each subset bivariate distribution to the global bivariate distribution.

The new global X marginal distribution, $f(x)$, has the following summary statistics:

$$\begin{aligned}
E\{X\} &= \sum_{i=1}^{m} p_i \cdot E\{X_i\} && \text{(B.3)}\\
E\{X^2\} &= \sum_{i=1}^{m} p_i \cdot (\sigma_i^2 + E\{X_i\}^2) && \text{(B.4)}\\
Var\{X\} &= E\{X^2\} - (E\{X\})^2 && \text{(B.5)}
\end{aligned}$$

Similarly, for the Y marginal distribution, $f_3(y)$,

$$\begin{aligned}
E\{Y\} &= \sum_{i=1}^{m} p_i \cdot E\{Y_i\} && \text{(B.6)}\\
E\{Y^2\} &= \sum_{i=1}^{m} p_i \cdot (\sigma_i^2 + E\{Y_i\}^2) && \text{(B.7)}\\
Var\{Y\} &= E\{Y^2\} - (E\{Y\})^2 && \text{(B.8)}
\end{aligned}$$

The resulting global bivariate distribution has the following covariance,

$$\begin{aligned}
E\{XY\} &= \sum_{i=1}^{m} p_i \cdot E\{X_iY_i\} && \text{(B.9)}\\
&= \sum_{i=1}^{m} p_i \cdot (Cov\{X_iY_i\} - E\{X_i\}E\{Y_i\}) &&\\
Cov\{XY\} &= E\{XY\} - E\{X\}E\{Y\} && \text{(B.10)}
\end{aligned}$$

$$\rho = \frac{Cov\{XY\}}{\sqrt{Var\{X\}Var\{Y\}}} \qquad \text{(B.11)}$$

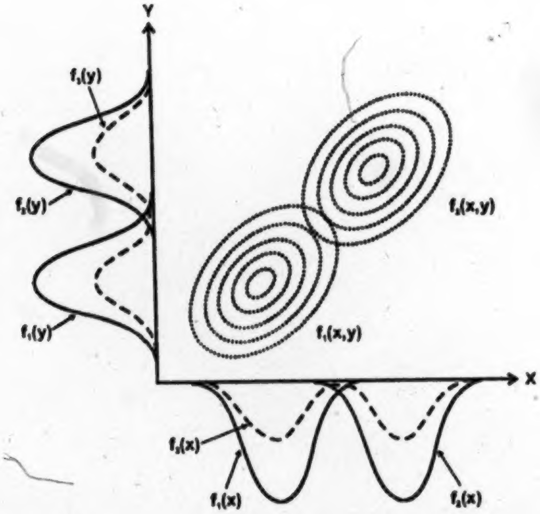


Figure B.3: Schematic illustration of combining subsets of bivariate distributions, $f_1(x, y)$ and $f_2(x, y)$. The marginal X and Y distributions corresponding to each bivariate distribution are given as $f_i(x)$ and $f_i(y)$, with i corresponding to that distribution. The result of combining these distributions are two marginal distributions, $f_3(x)$ and $f_3(y)$.

Figure B.3 is a schematic illustration for the case of $m = 2$; the resulting global bivariate distribution is bimodal and non-Gaussian.

Supposing that a global multivariate distribution can be constructed as a combination of conditional distributions, the decomposition of the global distribution to obtain a particular conditional distribution using the updating approach should be straightforward.

Proposed Methodology

The proposed algorithm for direct sequential cosimulation incorporates (1) the cokriging formalism for multiscale data, (2) a conversion from conditional means and variances to conditional distributions, (3) an iterative updating technique to obtain the conditional multivariate distribution, and (4) the stepwise decomposition of this distribution for cosimulation of multiscale data. Specifically, the main steps of the sequential algorithm are:

1. Pick a random path visiting all locations.
2. At each location:
    (a) Search for all nearby data of different types and/or scale and previously simulated nodes.
    (b) Perform simultaneous cokriging (collocated or full) to determine the parameters corresponding to the conditional univariate distribution for each variable.

(c) Using the cokriged parameters, determine the conditional univariate distribution for each variable using the approach proposed by Deutsch et. al. [18, 59]. These distributions will be referred to as the non-standard marginal distributions.

(d) Determine a non-standard multivariate distribution via the iterative updating approach:

$$f_{Y_1',Y_2'}(y_1, y_2) = f_{Y_2,Y_1}(y_2, y_1) \cdot \frac{f_{Y_1'}(y_1)}{f_{Y_1}(y_1)} \cdot \frac{f_{Y_2'}(y_2)}{f_{Y_2}(y_2)}$$

Resetting the global univariate and bivariate distributions to the previously updated distributions allows for iterative updating until the desired conditional univariate distributions are reproduced.

(e) Draw from the non-standard multivariate distribution in a stepwise manner:

  i. Draw a simulated value y_1 from the conditional marginal distribution of $Y_1(y_1)$.

  ii. From the conditional multivariate distribution determined in Step 2d, determine the conditional univariate distribution of $Y_2(y_2)$ given $Y_1 = y_1$, $f_{Y_2'|Y_1'} = y_1$. Draw y_2 from this conditional marginal distribution.

  iii. Repeat Step (ii) until a simulated value for each p variable is drawn.

(f) Proceed to next node.

Future Work

Naturally, implementation of the proposed methodology is a priority. It will be interesting to see how this procedure performs when it is applied to real, complex, multivariate data. The scaling approach should produce conditional multivariate distributions that respect the features inherent in the global multivariate distribution. These features may include non-linearity, constraints, and/or heteroscedasticity.

This will also serve to illustrate the effect of *not* reproducing the conditional correlations exactly. Of course, with the use of real data, the reference conditional distributions are not known. The only real comparison that will be possible is to determine whether the global multivariate distribution is reproduced. Should these results be encouraging, further theoretical and practical issues associated with the implementation will be explored. Simulation results will also be compared to those produced by some of the more conventional multivariate simulation techniques.

Future research in determining a multivariate conditional distribution will continue. The consequences of relying on a simple assumption of a stationary ratio (as in the proposed updating equation) will be explored. Future work will involve identification, exploration and validation of other methodologies that may be applied in order to achieve this same objective.